Markov invariants, plethysms, and phylogenetics[*]

J G Sumner[1,2], M A Charleston[1,4,5], L S Jermiin[3,4,5], and P D Jarvis[2,†]

[1] *School of Information Technologies,* [3] *School of Biological Sciences,* [4] *Centre for Mathematical Biology,* [5] *Sydney Bioinformatics, University of Sydney, NSW 2006, Australia*
[2] *School of Mathematics and Physics, University of Tasmania, TAS 7001, Australia*

Abstract
We explore model-based techniques of phylogenetic tree inference exercising Markov invariants. Markov invariants are group invariant polynomials and are distinct from what is known in the literature as phylogenetic invariants, although we establish a commonality in some special cases. We show that the simplest Markov invariant forms the foundation of the Log-Det distance measure. We take as our primary tool group representation theory, and show that it provides a general framework for analyzing Markov processes on trees. From this algebraic perspective, the inherent symmetries of these processes become apparent, and focusing on plethysms, we are able to define Markov invariants and give existence proofs. We give an explicit technique for constructing the invariants, valid for any number of character states and taxa. For phylogenetic trees with three and four leaves, we demonstrate that the corresponding Markov invariants can be fruitfully exploited in applied phylogenetic studies.

[*] This is the "long version" that includes an extended introduction, a subsection on mixed-weight invariants, a third appendix on the K3ST model, and a more relaxed pace with additional discussion throughout. The "short version" appears in *Journal of Theoretical Biology*, 253:601-615, 2008.

[†] Alexander von Humboldt Fellow

keywords: invariants, plethysm, phylogenetics, Schur functions, branching rules

email: jsumner@it.usyd.edu.au

UTAS-PHYS-2007-31

Contents

1 Introduction

1.1 Background

Molecular phylogenetic methods aim to infer the past evolutionary relationships of organisms from present day molecular data such as nucleotide sequences. Progress is made by making astute assumptions about the evolutionary process, which simplify the problem into a mathematical form, while retaining much of the structure motivating the biological question at hand. This process of mathematical modeling is essential if informed inferences from observed data sets are to be made.

The most significant simplification made in phylogenetic models is that the evolutionary change of the molecular units is assumed to progress by mutation under environmental influences and the Darwinian effects of selection are ignored. Another overriding simplification, featuring in all the popular models, is that the effect of mutations is modelled as a stochastic (random) process assumed to be Markov. Also, it is often assumed that any given site in a molecular sequence evolved independently of the other sites, and the probability of mutation at each site is identically distributed (known together as the IID assumption). Although the IID assumption is known *not* to hold in many cases [58], we will assume throughout that IID holds, and defer modification of the results presented here to this more general case.

Much progress in phylogenetic inference has been achieved in recent years with the use of sophisticated mathematics, probability and statistical theory, and the advent of powerful computing techniques. A general rule that rates the scientific credence of a phylogenetic method is that *model based* techniques are preferred. In particular, some recent work has focused on the elucidation of the implicit model assumptions of popular methods such as Neighbor-Joining [9, 26] and Maximum Parsimony [85]. This type of analysis is an essential part of the scientific justification because otherwise it is not exactly clear what is being estimated in the statistical sense. Without such a framework the biologist is left without any information regarding the confidence in the inference produced.

An overlying difficulty in phylogenetic tree inference is that the number of possible trees is vast, and the space of trees is non-Euclidean; hence it is not clear how one should proceed in searching through it. It is normal to begin with a candidate tree and then consider each of its neighbouring trees (under a given adjacency rule) and choose the new tree as that with the best score. There is a range of available tree perturbation types such as "prune and regraft" or Nearest Neighbour Interchange, which define these adjacencies. Which type is preferable is a matter of ongoing debate [14, 34] and such heuristic techniques sometimes find only locally optimal solutions. In this paper we will not discuss the problems associated with large trees, but consider how small trees may be built under general model assumptions. We give a general framework for constructing small trees, which can then be used as a springboard for building larger trees using techniques such as 'quartet puzzling' [78] or supertree methods (for arbitrarily sized subtrees) [8, 89].

Due to its importance for calculating divergence times of lineages, the rate of mutation present in models of evolution is of central importance in phylogenetics. There are several well-known limitations of the standard models involving the rate of mutation on a phylogenetic tree. For instance, the IID assumption is almost always violated by the existence of site-to-site rate variation [67] and by the existence of invariable sites [56]. Other issues include non-stationary processes of evolution (which leads to 'compositional heterogeneity' [44]), 'pattern heterogeneity' (where the pattern of substitutions varies across the sites [67]) and 'heterotachy' (differential rates across the tree) [57]. Ignoring the invalidity of the simple models when such assumptions are violated leads to model mis-specification [75] and (potentially) incorrect tree inference.

An issue for any inference technique is that of 'consistency'; where consistency is always with respect to an explicit or implicit model (or family of models) of sequence evolution. Statistical consistency requires that if the data set is sampled from a distribution generated under the model assumptions, then the inference method tends to the correct answer 100% of the time as the size of the data set (length of the sequences) tends to infinity. For example Felsenstein [21] showed that Maximum Parsimony (MP) is statistically *inconsistent* (with all but a small family of models [74]).

As exemplified by the first three chapters of the recent review book [25], the statistically consistent, model based phylogenetic methods can be placed into three categories: Minimum Evolution (ME) and distance based methods, Maximum Likelihood (ML), and Bayesian methods. ME proceeds by defining a (model based) matrix of pairwise distances between the molecular sequences, and then minimizes the total tree length across the space of possible trees subject to some statistical criteria such as least squares (see Chapter 1 of [25]). ML proceeds by maximizing the 'likelihood' of the observed data set across the set of possible trees and models of evolution [22, 25]. Bayesian methods proceed using Bayes' theorem to calculate a posterior distribution on the space of possible trees given a prior distribution (usually uniform–which is an issue in itself as this does not correspond to any evolutionary model of tree generation [93]). For each of these methods the underlying model assumptions are explicit, and current research efforts revolve around implementing these methods under expanded assumptions and/or in a computationally efficient manner.

Another desirable feature of any phylogenetic method is that the model on which it is based should be defined by as few numerical parameters as possible. The issue of scientific content of a model and parameter counts is discussed by Steel [75] in relation to the effectiveness of MP vs ML, where it was stated that the "predictive power of the theory. . . tends to be drowned out in a sea of parameter estimation". This is a fundamental problem in model selection for biological inference, and corresponds to what is known as the bias/variance trade off of parameter estimation [11]; which in turn equates to the problem of "overfitting" or "underfitting" a data set. From an information theoretic perspective, a given data set contains only so much information from which the numerical parameters of a model may be estimated. A model with many parameters may fit the data very well, in that the parameter estimates may be close to their true values, but the corresponding variances will be large because there are relatively few data points. On the other hand, the variance of the estimates of a model with very few parameters will be smaller as there are many data points to estimate each parameter, but in this case the model runs the risk of being badly mis-specified, so that the parameter estimates may be biased. In this light, the ‘covarion’ model [69] deals with the effects of invariable sites whilst introducing only one extra parameter, and the ‘gamma’ model [92] accounts for site-to-site rate variation, with only an additional two parameters. Other methods for coping with heterotachy, rate variation and pattern heterogeneity include the partitioning of data sets and mixture models [67]. However, all of these methods suffer because, in the general case, the models must include an individual rate matrix (containing up to twelve parameters) and an edge length parameter for each and every edge of the phylogenetic tree. In [70] it was recently noted that the task of phylogenetic tree inference often lies in a region where there are more parameters than data points.

To reduce the number of parameters in phylogenetic models, the evolutionary process is usually assumed to be stationary and reversible, the rate matrices are assumed to have a certain form (such as the Jukes-Cantor model with one parameter, or the Kimura models with two or three parameters), and each edge of the phylogenetic tree is assigned the same rate matrix (for details on these assumptions, see [10, 45]). To accommodate non-stationary processes and associated compositional heterogeneity, it becomes necessary to introduce many more parameters into the model. In this circumstance it then becomes desirable to use a technique based on a general model but without the need to estimate the numerical parameters. In this light, a matrix of Log-Det pairwise distances combined with the Neighbor-Joining algorithm [59] achieves statistically consistent tree inference under the assumption of a general model. However this technique has its own shortcomings as distance methods only consider pairwise sequence alignments, ignoring much of the information available in the data set, and has problems with model mis-specification [84], and the statistical properties of the Log-Det are not exactly known [29]. A recently presented method [42] fits a very general model, but clearly will have issues with over-parameterization and computational requirements.

In summary, the desirable features of a given phylogenetic method are that it is based on a general model of sequence evolution, it is statistically consistent with a family of known models, and the number of parameters to be estimated is minimal.

1.2 Markov invariants

In this work we introduce the use of mathematical representation theory to the problem of phylogenetic inference (further background to the results is presented in the PhD thesis [81]). We define ‘Markov invariants’ and show that these functions, when evaluated on sequence data, can be put to work in the problem of phylogenetic tree inference under rather general model assumptions.

Markov invariants are distinct from what is known in the literature as ‘phylogenetic invariants’ [13, 19, 51, 77]. Markov invariants are a particular case of group invariant functions [66] and are hence more constrained by definition than phylogenetic invariants. Some Markov invariants are simultaneously phylogenetic invariants, but the reverse is not true in general. The structure of Markov invariants is more akin to that of the Log-Det function [52, 59], which is constructed using the simplest example of a Markov invariant, yet it is not a phylogenetic invariant.

The appeal of this approach is that Markov invariants do not assume any particular rate matrices or edge length parameters on the phylogenetic tree. Broad conditions of molecular evolution are thus accommodated, incorporating arbitrary substitution rates, non-stationary and time-inhomogeneous processes, heterotachy, and arbitrary pattern heterogeneity across the tree. Further, Markov invariants satisfy certain algebraic relations for particular phylogenetic trees, and can provide a novel method of tree inference.

This approach to phylogenetic tree inference satisfies the desirable features given in the summary above. That is, Markov invariants are valid for a general model of sequence evolution, statistical consistency is assured, and only a few parameters need to be estimated.

In particular, for the quartet case, we give a tree inference routine, valid for these inclusive conditions, optimizing over only one parameter. It is hoped that, with additional understanding, this technique can be extended to larger trees. This will result in phylogenetic tree inference methods, valid for general models, that make use of only a few parameters. Such a possibility is very attractive, as all of the data is utilized, and a general model may be assumed with the risk of overfitting significantly reduced.

In this paper we outline the theoretical background required to understand the derivation of Markov invariants. This will necessitate, in §2, an excursion into elementary measure theory on finite sets, and the construction of 'phylogenetic tensors'. In §3 we analyse certain groups affiliated with the Markov process, and review standard results from group representation theory. This section concludes with a derivation of existence conditions for Markov invariants. In §4 we report on the structure of Markov invariants for phylogenetic trees with three and four leaves, and give examples of how they can be incorporated into practical phylogenetic analyses.

2 Measure theory, the Markov semigroup, and phylogenetic tensors

In §2.1 we collect some basic properties of measures on finite sets, justifying the use of tensor product spaces in the context of Markov processes on phylogenetic trees. The results are rather elementary, but ultimately necessary to place the subsequent discussion on its proper footing. See, for example, [30] for an introduction to measure theory. In §2.2 we use generating function techniques to calculate expectation values of various random variables (and functions thereof) associated with phylogenetic data sets. We give a simple example and show how to compute its unbiased estimator. We define the 'Markov semigroup' for the general time-inhomogeneous process (§2.3), construct 'phylogenetic tensors' (§2.4), and, finally, define Markov invariants (§2.5).

2.1 Probability measures on finite sets

Consider a finite set labelled by natural numbers, $K=\{1,2,\ldots,k\}$. A *probability measure* on K, is a function $\mu\colon K\rightarrow[0,1]$, such that, for any proper subset $A\subset K$ and any sequence $A_1,A_2,\ldots$ of pairwise disjoint subsets, the following conditions hold:

$$\begin{aligned}\mu(\varnothing)&=0,\\ \mu(A)&<1,\\ \mu\left(\bigcup_i A_i\right)&=\sum_i\mu(A_i),\\ \mu(K)&=1.\end{aligned}$$

We denote the set of probability measures on K as $\mathcal{M}(K)$. It follows from the third condition that for $1\leq i\leq k$ the measures, $\delta_i(A)=1$ if $i\in A$ and 0 otherwise, form a basis such that

$$\mu=\sum_{i=1}^{k}\mu_i\delta_i,$$

for all $\mu\in\mathcal{M}(K)$ with $\mu_i:=\mu\left(\{i\}\right)$. This definition is equivalent to the usual requirement of a probability distribution on a finite set:

$$\sum_{i=1}^{k}\mu_i=\mu\left(\bigcup_{i=1}^{k}\{i\}\right)=\mu(K)=1.$$

In phylogenetics the data sets under consideration are aligned sequences of molecular units. For example, in the case of DNA made up of the four nucleotides adenine, cytosine, guanine, thymine, we would have $K = \{A, C, G, T\}$, $k = 4$ and write $K = \{1, 2, 3, 4\}$. However, the results presented here and in §3 are valid for *any* k. In §4 we will concentrate on cases relevant to phylogenetics and investigate the Markov invariants for $k = 2, 3$ and 4.

In this work we do not consider the problem of aligning the sequence data, and assume throughout that the 'true' alignment (without gaps) can and has been found (where truth is relative to the modelling process). Under this circumstance, it becomes necessary to consider the direct product of K with itself m times:

$$K^m := \times^m K = K \times K \times \ldots \times K$$

with $|K^m| = k^m$. Exactly as above, for any proper subset $E \subset K^m$ and any sequence of pairwise disjoint subsets $E_1, E_2, \ldots$, a probability measure, $\mu \in \mathcal{M}(K^m)$, must equivalently satisfy

$$\begin{aligned} \mu(\varnothing) &= 0, \\ \mu(E) &< 1, \\ \mu\left(\bigcup_i E_i\right) &= \sum_i \mu(E_i), \\ \mu(K^m) &= 1. \end{aligned}$$

Given that under a measure unions decompose into summations, it follows that we have the tensor product:

$$\mathcal{M}(K^m) = \otimes^m \mathcal{M}(K) := \mathcal{M}(K) \otimes \mathcal{M}(K) \otimes \ldots \otimes \mathcal{M}(K).$$

Concretely, any subset of K^m can be expressed as a union of disjoint subsets of the form

$$A_1 \times A_2 \times \ldots \times A_m,$$

with $A_1, A_2, \ldots, A_m \subseteq K$. A basis for $\otimes^m \mathcal{M}(K)$ is then, for $1 \leq i_1, i_2, \ldots, i_m \leq k$,

$$\delta_{i_1} \otimes \delta_{i_2} \otimes \ldots \otimes \delta_{i_m}(A_1 \times A_2 \times \ldots \times A_m) := \delta_{i_1}(A_1)\delta_{i_2}(A_2)\ldots\delta_{i_m}(A_m),$$

with $\delta_{i_1}(A_1)\delta_{i_2}(A_2)\ldots\delta_{i_m}(A_m) = 1$ if $\{i_1\}\times\{i_2\}\times\ldots\times\{i_m\} \in A_1 \times A_2 \times \ldots \times A_m$ and 0 otherwise. We index the elements $\{i_1\}\times\{i_2\}\times\ldots\times\{i_m\}$ as

$$I = i_1 i_2 \ldots i_m,$$

and write

$$\mu_I \equiv \mu_{i_1 i_2 \ldots i_m} := \mu(\{i_1\}\times\{i_2\}\times\ldots\times\{i_m\}).$$

We refer to m as the *rank* of the tensor μ.

Previously the authors JGS and PDJ have presented probability distributions on phylogenetic trees in a tensor product formalism motivated from analogies to quantum physics [40, 83]. The formulation presented above places this construction on its proper measure-theoretic footing[1]. In §2.4 we will relate a given (Markov) model of evolution on a phylogenetic tree with m leaves, to a unique rank m tensor $P \in \otimes^m \mathcal{M}(K)$.

2.2 Random variables, generating function, expectation values and estimators

Any data set considered in a phylogenetic study is necessarily of finite extent, and we suppose that it is a sample drawn from some unknown distribution. We wish to define expectation values of such data (or events) and functions thereof. Throughout we will assume the IID assumption

[1] We are indebted to Michael Baake for drawing our attention to this.

holds, so that we need only consider the distribution of a single random variable. The probability of observing a particular state at a given site will be identical for all the other sites.

For a set of m aligned sequences of length N, define a *pattern* to be the (ordered) set of states read across the m sequences at a particular site in the alignment. That is, a pattern takes the form $I = i_1 i_2 \dots i_m$, where i_a is the character state in the a^{th} sequence. Define the random variable X as the pattern observed at a given site. A probability distribution for X can be specified using a probability measure $\mu \in \otimes^m \mathcal{M}(K)$:

$$\mathbb{P}[X = i_1 i_2 \dots i_m] = \mu_{i_1 i_2 \dots i_m}. \tag{1}$$

For a sequence of finite length N, define Z as the random variable that counts the number of occurrences of each pattern $I = i_1 i_2 \dots i_m$ in the alignment, so that

$$Z = (Z_I) = (Z_{i_1 i_2 \dots i_m})_{1 \leq i_1, i_2, \dots, i_m \leq k},$$

and $\sum_{I \in K^m} Z_I = N$. Assuming that each site in the alignment is identically and independently distributed as (1), it follows that Z is multinomially distributed under the measure μ:

$$\mathbb{P}[Z = z; N] = \prod_{I \in K^m} \frac{N!}{z_I!} \mu_I^{z_I}.$$

This expresses, under the assumptions of μ, the probability of observing within the alignment of m sequences the specific number of occurrences of each of the possible character patterns $Z = z$.

When we describe Markov invariants, we will need to discuss expectation values of the random variable Z and functions thereof. For any function ϕ, the expectation value with respect to the measure μ is defined as

$$E[\phi(Z)] := \sum_z \phi(z) \mathbb{P}[Z = z; N],$$

with the summation over all z such that $\sum_{I \in K^m} z_I = N$.

Remembering that Z follows a multinomial distribution, it is in practice necessary to use generating function techniques in order to calculate these expectation values. The generating function on the formal variables $s = (s_I) = (s_{i_1 i_2 \dots i_m})_{1 \leq i_1, i_2, \dots, i_m \leq k}$ of the multinomial distribution is

$$G(s) := E[e^{(s,Z)}] = \left(\sum_{I \in K^m} \mu_I e^{s_I} \right)^N, \tag{2}$$

with

$$(s, Z) := \sum_{I \in K^m} s_I Z_I.$$

From the properties of the exponential function and the commutivity of differentiation and expectation,

$$\left. \frac{\partial G(s)}{\partial s_{i_1 i_2 \dots i_m}} \right|_{s=0} = E[Z_{i_1 i_2 \dots i_m}].$$

Using the above closed form of the generating function, an elementary calculation returns

$$E[Z_{i_1 i_2 \dots i_m}] = N \mu_{i_1 i_2 \dots i_m},$$

as of course would be expected. This can be extended to find the expectation of any function of Z:

$$E[\phi(Z)] = \phi \left(\frac{\partial}{\partial s} \right) G(s) \Big|_{s=0}.$$

As a concrete example, take $m = 2$ and consider the case $\phi(Z) = Z_{44}^2 - Z_{12}Z_{13}$. From the linearity of the expectation values we have

$$E[Z_{44}^2 - Z_{12}Z_{13}] = E[Z_{44}^2] - E[Z_{12}Z_{13}],$$

so we can consider each term in turn. Taking derivatives of the closed form of the generating function gives

$$E[Z_{44}^2] = N(N-1)\mu_{44}^2 + N\mu_{44}$$

and

$$E[Z_{12}Z_{13}] = N(N-1)\mu_{12}\mu_{13}.$$

Thus, in this case, the expectation value of ϕ is

$$E[\phi(Z)] = N(N-1)(\mu_{44}^2 - \mu_{12}\mu_{13}) + N\mu_{44}.$$

Given a (possibly unobservable) random variable θ, an *estimator* is another random variable which is a function of observable quantities such that its expectation value somehow approximates θ. The *bias* of an estimator $\widehat{\theta}$ is defined as the difference

$$b(\hat{\theta}) = E[\widehat{\theta}] - E[\theta],$$

allowing for θ to simply be a constant so that $E[\theta] = \theta$. An *unbiased* estimator is simply an estimator with bias equal to zero. For example, a short calculation reveals that the unbiased estimator of $\phi(\mu)$ above is

$$\frac{\phi(Z) - Z_{44}}{N(N-1)}.$$

In general, if ϕ is polynomial, computing an unbiased form is a straightforward matter of solving a sequence of difference equations. When it comes to discussing estimators for Markov invariants, we will show that unbiased forms can easily be defined. However, we will note that explicit computation is difficult due to a required change of basis.

2.3 The Markov semigroup

A stochastic process can be described by introducing a time-dependent random variable $X(t)$. A crucial component of the subsequent discussion will be that the time evolution of the corresponding probability distribution can be viewed as a linear mapping upon a vector space. Presently we will establish the conditions for a *Markov* process, and show that such a process satisfies the desired property. See, for example, [38] for an equivalent derivation.

Consider a time-dependent, finite-state random variable, $X(t)$, taking on values in K, any set of times $t_1 < t_2 < \ldots < t_n < t$, and the joint distribution of X across those times:

$$\mathbb{P}[X(t_1)=i_1, X(t_2)=i_2, \ldots, X(t_n)=i_n, X(t)=i].$$

The distribution of X at the particular time t is given by the marginal,

$$\mathbb{P}[X(t)=i] = \sum_{1\leq i_1,i_2,\ldots,i_n\leq k} \mathbb{P}[X(t_1)=i_1, X(t_2)=i_2, \ldots, X(t_n)=i_n, X(t)=i],$$

and this can be re-expressed by invoking the conditional distribution:

$$\mathbb{P}[X(t)=i] = \sum_{1\leq i_1,i_2,\ldots,i_n\leq k} \mathbb{P}[X(t)=i|X(t_1) = i_1, X(t_2)=i_2, \ldots, X(t_n)=i_n] \cdot \mathbb{P}[X(t_1)=i_1, X(t_2)=i_2, \ldots, X(t_n)=i].$$

The simplest stochastic process is the process for which the probability of a transition to a new state at a given time is independent of the states at all preceding times (such as tossing of a coin–the *Bernoulli process*). A *Markov process* can be seen as the next simplest case where the probability of a transition is independent of all but the state at the most recent time. Thus, for a Markov process the conditional distribution satisfies

$$\mathbb{P}[X(t)=i|X(t_1)=i_1, X(t_2)=i_2, \ldots, X(t_n)=i_n] = \mathbb{P}[X(t)=i|X(t_n)=i_n].$$

This implies that the marginal distribution of X at the time t is

$$\begin{aligned}\mathbb{P}[X(t)=i] = &\sum_{1\leq i_n\leq k} \mathbb{P}[X(t)=i|X(t_n)=i_n] \\ &\qquad\cdot \sum_{1\leq i_1,\ldots,i_{n-1}\leq k} \mathbb{P}[X(t_1)=i_1, X(t_2)=i_2, \ldots, X(t_n)=i_n] \\ = &\sum_{1\leq i_n\leq k} \mathbb{P}[X(t)=i|X(t_n)=i_n]\mathbb{P}[X(t_n)=i_n].\end{aligned}$$

Introducing the time-dependent measure μ^t with $\mu^t(\{i\}){:=}\mu^t_i{=}\mathbb{P}[X(t){=}i]$, we can express this as

$$\mu^t_i = \sum_{1\leq j\leq k} M_{ij}(t,s)\mu^s_j,$$

for all $s<t$, and for $M_{ij}(t,s){:=}\ \mathbb{P}[X(t){=}i|X(s){=}j]$. If we consider the $(M_{ij}(t,s))_{1\leq i,j\leq k}$ as the matrix elements of a linear operator $M(t,s)$ acting on the vector space $\mathbb{R}^k \supset \mathcal{M}(K)$ with basis elements $\delta_1, \delta_2, \ldots, \delta_k$, we see that, as promised, for a Markov process the time evolution of the probability distribution is given by a linear map on $\mathbb{R}^k$ defined by its action on time-dependent probability measures:

$$\begin{aligned}&\mu^s \overset{M(t,s)}{\mapsto} \mu^t, \\ &\mu^t = M(t,s)\mu^s.\end{aligned} \tag{3}$$

This linear map describes the general time-inhomogeneous finite state Markov process and can easily be extended to the whole of $\mathbb{R}^k$.

In [83] JGS and PDJ considered stochastic matrices as elements of the general linear group, and used this property to study the structure of invariant polynomials (used as measures of entanglement in quantum physics) when evaluated on a phylogenetic tree. Presently we will define the Markov semigroup which serves to refine the definition of invariant functions to the more relevant case of a stochastic (but linear) time evolution.

Define the time-dependent *rate matrix*, $Q(t)$, as a (continuous) one-parameter family of linear operators on the vector space $\mathcal{M}(K)$, which in the $\delta_1, \delta_2, \ldots, \delta_k$ basis has matrix elements satisfying:

$$Q_{ij}(t) \geq 0, \quad \forall i \neq j; \qquad Q_{ii}(t) = -\sum_{j\neq i} Q_{ji}(t).$$

The summation conditions can be equivalently expressed by defining the vector $\theta = \delta_1+\delta_2+\ldots+\delta_k$ and its transpose $\theta^\top$, and setting

$$\theta^\top Q(t) = 0,$$

for all t.

The Markov semigroup on k elements, $\mathfrak{M}(k)$, with parameters $0\leq s\leq t<\infty$, is defined as the subset of (differentiable) two-parameter linear operators on $\mathcal{M}(K)$ which satisfy

$$M(t,s) = 1, \qquad \forall t = s;$$

the Chapman-Kolmogorov equation:

$$M(t,s)M(s,r) = M(t,r), \qquad \forall r < s;$$

and the backwards and forwards equations:

$$\begin{aligned} \frac{\partial M(t,s)}{\partial s} &= -M(t,s)Q(s), \\ \frac{\partial M(t,s)}{\partial t} &= Q(t)M(t,s); \end{aligned} \tag{4}$$

for any rate matrix $Q(t)$ [27, 38]. Solutions of (4) can be represented using the time-ordered product (or ordered-exponential):

$$M(t,s) = \mathbb{T}\exp\int_s^t Q(u)du \tag{5}$$

[39, Chap. 4], from which it follows that

$$\det M(t,s) = \exp\int_s^t tr(Q(u))du, \tag{6}$$

and

$$\theta^\top M(t,s) = \theta^\top.$$

The time-ordered product is best understood by considering the approximation

$$M(s+2\epsilon,s) = M(s+2\epsilon,s+\epsilon)M(s+\epsilon,s) \simeq e^{Q(s+\epsilon)\epsilon}e^{Q(s)\epsilon}.$$

By considering (4) for the case $t = s$, it follows that in the $\delta_1, \delta_2, \ldots, \delta_k$ basis, the matrix elements of each $M(t,s)$ lie in the interval $[0,1]$ for all $s \leq t$. Thus, the Markov semigroup corresponds to the subset of the set of stochastic matrices subject to the condition that for each matrix there exists a rate matrix (or generator) $Q(t)$ such that (5) is satisfied. We refer to elements of the Markov semigroup as *Markov operators*.

In the time-homogeneous case where the rate matrix is time-independent:

$$Q := Q(t) = Q(0),$$

it follows that $M(t,s)$ is dependent only upon the difference $(t-s)$, and form (5) becomes simply

$$M(t) = e^{tQ} = \sum_{0\leq n<\infty} \frac{(tQ)^n}{n!}.$$

In §3 we will discuss some representation-theoretic properties of certain groups affiliated with the Markov semigroup.

2.4 Phylogenetic tensors

A tree, $\mathcal{T}$, is a connected graph without cycles and consists of a set of vertices and edges. Vertices of degree one are called *leaves*. We work with *oriented* trees, which are defined by directing each edge of $\mathcal{T}$ away from a distinguished vertex, ρ, known as the *root* of the tree. Consequently, a given edge lying between adjacent vertices u and v is specified as an ordered pair (u,v), where u lies on the unique path from ρ to v. A *cherry* is a pair of leaf vertices with the same parent vertex.

Assign a random variable, X_v, to each vertex of the tree, and, as described in [72, Chap. 8], a joint distribution of the random variables at the leaves is determined by specifying a distribution $\pi \in \mathcal{M}(K)$ at ρ and a Markov operator $M^{v,u} \in \mathfrak{M}(k)$ for every edge (u,v). In particular, for every

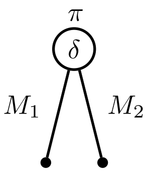

Figure 1: Phylogenetic tree with two leaves

v, the random variable X_v is conditional on only the random variables lying on the path from ρ and v, and for each pair of vertices v_1, v_2 with common parent u, the joint distribution of X_{v_1} and X_{v_2} is given by

$$\mathbb{P}[X_{v_1}\!=\!i_1, X_{v_2}\!=\!i_2] = \sum_{1\leq j\leq k} M^{v_1,u}_{i_1 j} M^{v_2,u}_{i_2 j} \mu^u_j, \tag{7}$$

where μ^u is the distribution of X_u. The empirical interpretation of the joint distribution across the leaves is that of a sampling distribution from which an alignment of molecular sequences is constructed by drawing one character pattern at a time. Throughout this paper, we will consider phylogenetic trees where the root distribution and the Markov operators are arbitrary.

Note that we insist that the Markov operators belong to the Markov semigroup, so a continuous-time process is in action throughout the tree. This additional analytic structure means that this model is slightly less general than the *general Markov model* as defined in [1, 41]. The general Markov model allows for arbitrary transition matrices with positive entries and unit row-sum (unit column-sum in our formulation), and it is not hard to find a matrix satisfying these conditions but with determinant less than or equal to zero, directly contradicting (6).

We will consider a joint probability distribution on m leaves as a probability measure, $P \in \otimes^m \mathcal{M}(K)$, that we refer to as a *phylogenetic tensor*. Presently we review how these tensors can be constructed using purely algebraic operations.

The branching process (7) can be interpreted as a map that takes probability measures on K to probability measures on $K\!\times\! K\!=\!K^2$. In [40, 83], it was shown how to formalize this by defining the linear operator

$$\delta : \mathcal{M}(K) \rightarrow \mathcal{M}(K) \otimes \mathcal{M}(K).$$

Demanding the conditional dependencies that are required for the standard definition of a tree distribution [72, Chap. 8], we have (expressed in the $\delta_1, \delta_2, \ldots, \delta_k$ basis) the specification

$$\delta : \delta_i \mapsto \delta_i \otimes \delta_i, \qquad 1 \leq i \leq k.$$

The phylogenetic tree with two leaves (Figure 1) can then be represented as the string

$$P = (M_1 \otimes M_2) \cdot (\delta \cdot \pi),$$

where M_1 and M_2 are the Markov operators on the two edges of the tree and, if X_1 and X_2 are the random variables at the leaves 1 and 2, respectively, we have

$$\mathbb{P}[X_1\!=\!i, X_2\!=\!j] = P_{ij} := P(\{i\}\!\times\!\{j\}).$$

This construction can be generalized to any phylogenetic tree by colouring the root of the tree with a distribution π, each internal vertex (including the root) with a branching operator δ, and every edge with an arbitrary Markov operator. The phylogenetic tensor is constructed by beginning at the root of the tree, and then recursively moving to the child vertices and applying the relevant operators to the corresponding slots in the (growing) tensor. Whenever a leaf is encountered, continually apply the identity operator at that leaf, until all leaves have been reached and the phylogenetic tensor is complete. A phylogenetic tensor, P, is then represented as a string made

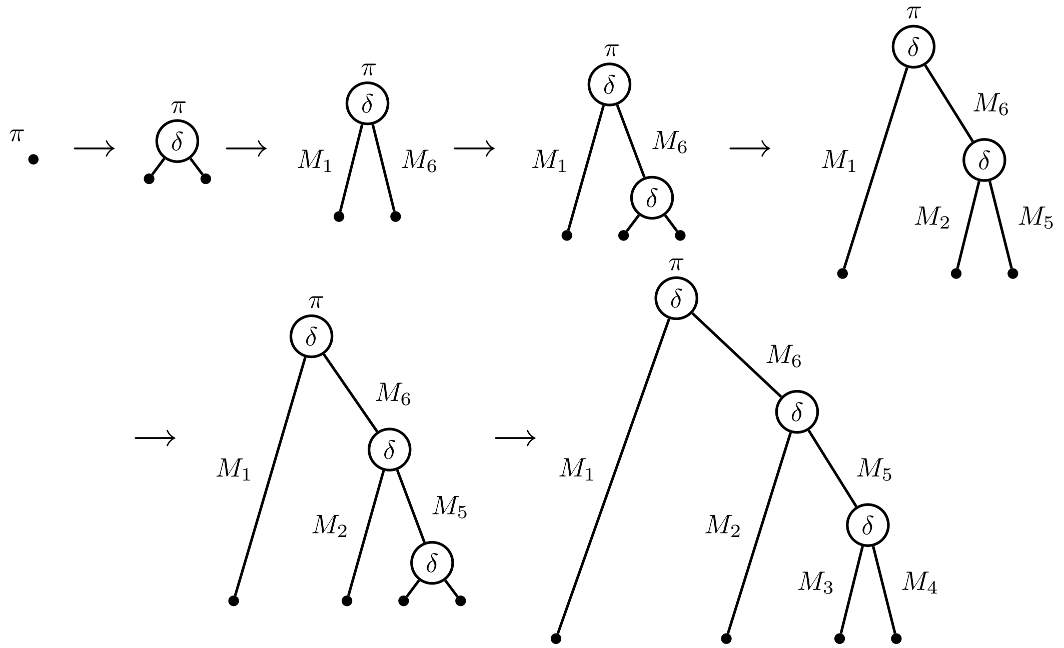

Figure 2: Constructing the phylogenetic tensor for a four taxon tree

up of the characters π, $M_1, M_2, \ldots,$ and δ, and the joint distribution of the random variables $X_1, X_2, \ldots, X_m$ at the leaves $1, 2, \ldots, m$ is given by

$$\mathbb{P}[X_1 = i_1, X_2 = i_2, \ldots, X_m = i_m] = P_{i_1 i_2 \ldots i_m} := P(\{i_1\} \times \{i_2\} \times \ldots \times \{i_m\}).$$

For example, the phylogenetic tensor of four leaves (Figure 2) is represented by the string

$$P = (1 \otimes 1 \otimes M_3 \otimes M_4) \cdot (1 \otimes 1 \otimes \delta) \cdot (1 \otimes M_2 \otimes M_5) \cdot (1 \otimes \delta) \cdot (M_1 \otimes M_6) \cdot (\delta \cdot \pi),$$

and is constructed in the steps

$$\begin{aligned}
&\pi \to \delta \cdot \pi \to (M_1 \otimes M_6) \cdot (\delta \cdot \pi) \to (1 \otimes \delta) \cdot (M_1 \otimes M_6) \cdot (\delta \cdot \pi) \\
&\quad \to (1 \otimes M_2 \otimes M_5) \cdot (1 \otimes \delta) \cdot (M_1 \otimes M_6) \cdot (\delta \cdot \pi) \\
&\quad\quad \to (1 \otimes 1 \otimes \delta) \cdot (1 \otimes M_2 \otimes M_5) \cdot (1 \otimes \delta) \cdot (M_1 \otimes M_6) \cdot (\delta \cdot \pi) \\
&\quad\quad\quad \to (1 \otimes 1 \otimes M_3 \otimes M_4) \cdot (1 \otimes 1 \otimes \delta) \cdot (1 \otimes M_2 \otimes M_5) \cdot (1 \otimes \delta) \cdot (M_1 \otimes M_6) \cdot (\delta \cdot \pi).
\end{aligned}$$

In order to define Markov invariants, we must also define two reduced tensors based on P, the *trimmed* tensor $\widetilde{P}$ and the *pruned* tensor P^*. These are both constructed by modifying the underlying tree. The trimmed tensor $\widetilde{P}$ is constructed by taking P and setting the Markov operators on the pendant edges all equal to the identity operator, or equivalently setting the lengths of the pendant edges to zero. The pruned tensor P^* is constructed by removing all cherries from the trimmed tensor. The rank of the pruned tensor is $(m - c)$ where c is the number of cherries on the underlying tree.

In the general case, we can relate P and $\widetilde{P}$ as

$$P = (M_1 \otimes M_2 \otimes \ldots \otimes M_m) \cdot \widetilde{P}, \tag{8}$$

where $M_1, M_2, \ldots, M_m$ are the Markov operators on the leaf edges. In what is to come, we will continually use this relation.

As an illustration of the relationship between P, $\widetilde{P}$ and P^*, take the seven leaf tree (Figure 3), with phylogenetic tensor given by

$$\begin{aligned}
P = (1 \otimes M_2 \otimes M_3 \otimes M_4 \otimes M_5 \otimes M_6 \otimes M_7) \cdot (1 \otimes \delta \otimes \delta \otimes \delta) \\
\cdot (M_1 \otimes M_8 \otimes M_9 \otimes M_{10}) \cdot (\delta \otimes \delta) \cdot (M_{11} \otimes M_{12}) \cdot (\delta \cdot \pi).
\end{aligned}$$

The trimmed tensor corresponding to the tree (Figure 4) is obtained by clipping off the pendant edges:

$$\widetilde{P}=(1\otimes\delta\otimes\delta\otimes\delta)\cdot(1\otimes M_8\otimes M_9\otimes M_{10})\cdot(\delta\otimes\delta)\cdot(M_{11}\otimes M_{12})\cdot(\delta\cdot\pi),$$

and, finally the pruned tensor corresponding to the tree (Figure 5) is expressed as:

$$P^*=(M_8\otimes M_9\otimes M_{10})\cdot(1\otimes\delta)\cdot(M_{11}\otimes M_{12})\cdot(\delta\cdot\pi).$$

2.5 Markov invariants, definition

With the form (8) in mind, we define a *Markov invariant* of *weight* $(w_1,w_2,\ldots,w_m)$ as a function satisfying

$$f(P)=(\det M_1)^{w_1}(\det M_2)^{w_2}\ldots(\det M_m)^{w_m}f(\widetilde{P}), \tag{9}$$

for all $M_1,M_2,\ldots,M_m\in\mathfrak{M}(k)$. We exclusively consider polynomial functions, and where $w_1=w_2=\ldots=w_m\equiv w$, the Markov invariant is said to be of weight w.

Considering the above discussion of unbiased estimators of random variables, an unbiased estimator of a Markov invariant is a function, $\widehat{f}$, such that

$$E[\widehat{f}(Z)]=f(P)=(\det M_1)^{w_1}(\det M_2)^{w_2}\ldots(\det M_m)^{w_m}f(\widetilde{P}).$$

Such an estimator depends, up to the multiplicative scaling factor, only upon the *internal* structure of the phylogenetic tree. It is exactly this property that can be productively engaged in the context of phylogenetic tree inference.

Conversely, a *phylogenetic invariant* is a function satisfying $f(P)\equiv 0$ for all P belonging to the family of phylogenetic tensors arising from a particular tree (or subset of trees). In §4 we will show that there exist Markov invariants for trees with three and four leaves that are simultaneously phylogenetic invariants.

Given a Markov invariant, f, consider the induced function, f^*, defined on pruned tensors and specified by evaluating the trimmed tensor:

$$f^*(P^*)=f(\widetilde{P}).$$

This induced function is easily extended to be defined upon all of $\otimes^{m-c}\mathcal{M}(K)$, where c is the number of cherries on the underlying tree of P. Such cases are of special interest for phylogenetic problems. In §4 we will review a case (reported in a less general context in [84]) where this induced function is itself a Markov invariant. We expect that future investigations of Markov invariants will reveal more cases such as this.

In §3.5 we will establish existence conditions for Markov invariants using standard results from group representation theory.

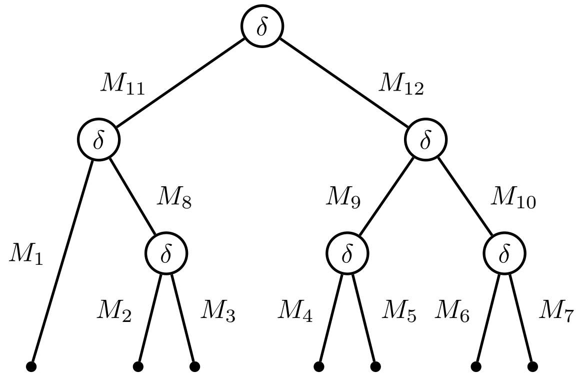

Figure 3: Phylogenetic tensor P for seven taxon tree ((1,23),(45,67))

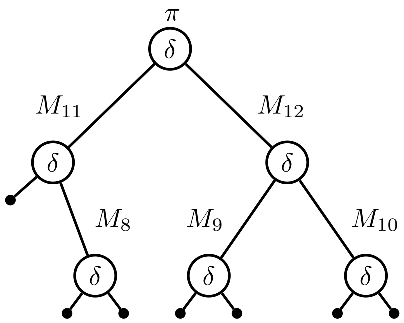

Figure 4: Trimmed phylogenetic tensor $\widetilde{P}$

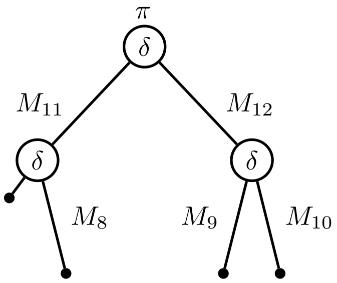

Figure 5: Pruned phylogenetic tensor P^*

3 Group representation theory in phylogenetics

In this section we use the algebraic description of probability distributions on phylogenetic trees given in §2.4 to establish natural connections with aspects of representation theory, for certain groups affiliated to the Markov semigroup. These are discussed in §3.1. Then follows (§3.2 and §3.3) a brief outline of those aspects of the representation theory of the general linear group and its subgroups that are needed for the discussion of group branching rules. This leads to the construction of one-dimensional representations and their identification as invariants (§3.4), with existence conditions given in §3.5.

3.1 The Markov semigroup and affiliated groups

The linear transformation (3) effected under the Markov semigroup on probability measures is closely related to certain group actions on the vector space $\mathbb{R}^k$. Given that the corresponding representation theory is unchanged [48], in this section we will generalize to complex vector space, (as with [1]). That is, here and below, for algebraic purposes we regard the $\delta_1, \delta_2, \ldots, \delta_k$ as elements of a basis for $V \cong \mathbb{C}^k$. Thus, the probability measures become a subset lying in the ambient complex space $\otimes^m \mathbb{C}^k \supset \otimes^m \mathcal{M}(K)$. For related considerations involving the study of invariants of stochastic matrices see [46, 65].

Referring to (6) and noting that $-\infty < tr(Q(t)) \leq 0$ for all t, the determinant of each element $M(t,s)$ lies in the interval $(0,1]$, and the Markov semigroup occurs as a subset of the general linear group:

$$\mathfrak{M}(k) \subset GL(k).$$

$GL(k)$ is the group of *invertible* linear operators on the k-dimensional vector space $\mathbb{C}^k$. The smallest subgroup of $GL(k)$ that contains $\mathfrak{M}(k)$ is obtained by taking $\mathfrak{M}(k)$ together with all of its operator inverses. In order to apply known methods of representation theory, we will, however, not work with this group directly. We define a slightly less refined subgroup as the focus of the impending discussion.

Generalizing the notation of [65], we define the subgroup $GL_1(k) \lhd GL(k)$ as the subset of $GL(k)$

whose matrices in the $\delta_1, \delta_2, \ldots, \delta_k$ basis have unit column-sum. That is, for all $g \in GL_1(k)$:

$$\theta^\top g = \theta^\top.$$

The group property clearly holds, as for all $g_1, g_2 \in GL_1(k)$:

$$\theta^\top(g_1 g_2) = (\theta^\top g_1) g_2 = \theta^\top g_2 = \theta^\top.$$

This group is isomorphic to the complex *affine group*[2]

$$GL(k-1) \ltimes T(k-1) \equiv A(k-1),$$

where $T(k-1)$ is the group of linear translations on $\mathbb{C}^{k-1}$. As shown in Appendix A, this isomorphism is due to the column-sum condition being, in effect, a statement that the group elements are dual to linear transformations in k-dimensional complex space, leaving a fixed vector invariant.

Consider also the *doubly-stochastic* Markov semigroup, $\mathfrak{M}^*(k)$, obtained by requiring an additional condition on the rate matrices:

$$Q(t)\theta = 0.$$

The associated subgroup of the general linear group is then denoted as $GL_{1,1}(k)$; the subgroup of matrices in $GL(k)$ which have unit column- *and* row-sum with, for all $g \in GL_{1,1}(k)$:

$$\begin{aligned} \theta^\top g &= \theta^\top, \\ g\theta &= \theta. \end{aligned}$$

Again the group property can easily be shown to hold. Thus the doubly-stochastic Markov semigroup is naturally affiliated to the associated group $GL_{1,1}(k)$ which, also as shown in Appendix A, itself is isomorphic to $GL(k-1)$.

To summarise, consider the subgroup chain:

$$GL(k-1) \cong GL_{1,1}(k) \lhd GL(k-1) \ltimes T(k-1) \equiv A(k) \cong GL_1(k) \lhd GL(k). \tag{10}$$

and the set inclusions:

$$\begin{aligned} \mathfrak{M}(k) &\subset GL_1(k), \\ \mathfrak{M}^*(k) &\subset GL_{1,1}(k). \end{aligned}$$

We now have a clear picture of how to develop the representation theory of the Markov semigroup which focuses on algebraic properties and avoids the analytic details due to the positivity requirement and semigroup property. This is the correct framework in which to exploit the Schur-Weyl duality (§3.2) and, considering the above inclusions, all results presented will be valid for the Markov semigroup. The above subgroup chain will feature in §3.4 where we give existence conditions for Markov invariants.

3.2 Representations of $GL(k)$ and Schur-Weyl duality

Our purpose here is to show that the close relation of the Markov semigroup to affiliated subgroups of the general linear group allows the machinery of representation theory to be applied in analysing the models used in phylogenetic inference.

From the form of the general Markov model on phylogenetic trees given earlier (8), it is evident that the representation-theoretic considerations must be extended to tensor products. We now provide some standard results within this setting (see, for example, [24, Lecture 6]).

[2] The symbol $\ltimes$ denotes the *semi-direct product* of two groups [3, Chap. 1]. The standard physical example is the Euclidean group, which occurs as the semi-direct product between rotations and translations in $\mathbb{R}^n$. These are none other than the set of transformations that define Euclidean geometry.

For $GL(k)$ and its classical subgroups it is well known that for the *defining* representation on $V \cong \mathbb{C}^k$, with $v \mapsto gv$, extended to a reducible representation on $\otimes^m V$ in the obvious way, $v_1 \otimes v_2 \otimes \ldots \otimes v_m \mapsto gv_1 \otimes gv_2 \otimes \ldots \otimes gv_m$, there is a direct sum decomposition,

$$\otimes^m V = \sum_{\lambda \vdash m} \oplus f_\lambda V^\lambda, \tag{11}$$

into (possibly reducible) subspaces V^λ. These subspaces (or *modules*) are labelled by integer partitions, $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$, of m, the λ_i being nonzero and nonincreasing and such that $\lambda_1 + \lambda_2 + \ldots + \lambda_n = m$. If λ is a partition of m, we write $\lambda \vdash m$ and $|\lambda| = m$. The corresponding module V^λ is determined by a unique projector on $\otimes^m V$; the *Young's operator* Y^λ. The f_λ are integer multiplicities determining how many times each module occurs in the decomposition. The *Schur-Weyl duality* is the classic result that each f_λ is none other than the dimension of the irreducible representation associated with the same partition λ of the symmetric group $\mathfrak{S}_m$. This reflects the role of the symmetric group's action on $\otimes^m V$ by permuting basis elements across the tensor product, when constructing the Young's operators.

The *character* of a representation is defined as the set of traces of the representing matrices; one for each group element. The irreducible representations of a group can be enumerated by solely considering the corresponding irreducible characters. Thus the problem of decomposing a representation into irreducible modules (computing the multiplicities f_λ) can be performed at the level of the characters[3].

For instance, in the case of $GL(k)$ itself, the V^λ are irreducible, with character given by the celebrated Schur functions, s_λ, with

$$s_\lambda(x) = \mathrm{tr}(\pi_\lambda(g)),$$

where $\pi_\lambda(g)$ is the representing matrix for group element g and $x_1, x_2, \ldots, x_k$ are its eigenvalues. The Schur functions are defined in their own right, and are uniquely determined by the semi-standard tableaux corresponding to the partition λ [60].

The defining k-dimensional representation in this notation is $V^{\{1\}} \cong \mathbb{C}^k$, in which case the Schur function is $s_{\{1\}}(x) = x_1 + x_2 + \ldots + x_k$. The Schur functions form a basis for the ring of symmetric functions on any number of variables, and the trace is a symmetric function. Hence, the problem of identifying the irreducible $GL(k)$ modules in the above representation on $\otimes^m V$, reduces to identifying the Schur functions in the decomposition of the character with respect to this basis[4].

A convenient and standard notation for Schur functions is given by enclosing the partition (or parts thereof) in braces [54]. Thus $\{\lambda\}$ and $\{1\}$, are the Schur functions corresponding to a general irreducible and the defining representation of $GL(k)$ respectively. For simplicity, we write $\pi_{\{1\}}(g) = g$.

For classical subgroups of $GL(k)$, the modules V^λ are no longer necessarily irreducible, and further combinatorial considerations (not required here) are needed to effect a complete reduction[5]. More importantly, for $GL(k)$ itself with V not the defining, but an arbitrary module, V^ρ say, the equivalents of the above modules, $(V^\rho)^\lambda$, are again no longer irreducible in general.

This construction introduces a fundamental operation for combining representations together; that of *plethysm* [54]. The character of $(V^\rho)^\lambda$ is denoted $\{\rho\}\underline{\otimes}\{\lambda\}$; the plethysm of $\{\rho\}$ by $\{\lambda\}$. In the simplest case $\{\rho\}$ is the character for the defining representation, $\{1\}$, and by definition $\{1\}\underline{\otimes}\{\lambda\} = \{\lambda\}$.

In general, for any symmetric functions A, B we have $\{\rho\}\underline{\otimes}(A + B) = \{\rho\}\underline{\otimes}A + \{\rho\}\underline{\otimes}B$, and

[3] Within the context of phylogenetics, see [62] for an unrelated discussion of the irreducible characters of the symmetric group.

[4] The stronger statement that the V^λ provide the complete set of irreducible modules of *any* integral representation of $GL(k)$ is valid [48].

[5] The classical subgroups of $GL(k)$ are constructed by requiring, under the group action, the invariance of bilinear forms on V.

we recover

$$\{\rho\}\underline{\otimes}\big(\sum_{\lambda\vdash m} f_\lambda\{\lambda\}\big) =\{\rho\}\otimes\{\rho\}\otimes\ldots\otimes\{\rho\},$$

where $\{\rho\}\otimes\{\lambda\}$ denotes the (commutative and associative) pointwise multiplication of the Schur functions,

$$(\{\rho\}\otimes\{\lambda\})(x) := \{\rho\}(x)\cdot\{\lambda\}(x),$$

and the Schur functions occurring in the decomposition of $\{\rho\}\otimes\{\lambda\}$ correspond to partitions of $|\rho|+|\lambda|$. This of course reflects (11) with V replaced by V^ρ:

$$V^\rho\otimes V^\rho\otimes\ldots\otimes V^\rho = \sum_{\lambda\vdash m}\oplus f_\lambda (V^\rho)^\lambda.$$

In particular, for rank 2 we have

$$V^\lambda\otimes V^\lambda = (V^\lambda)^{\{2\}}\oplus(V^\lambda)^{\{1^2\}},$$

which at the level of the characters is described completely by

$$\{\lambda\}\otimes\{\lambda\} = \{\lambda\}\underline{\otimes}\{2\}\,+\,\{\lambda\}\underline{\otimes}\{1^2\}.$$

This is the well-known decomposition of a representation into its symmetric and anti-symmetric Kronecker square, respectively[6].

Although it is a difficult task to evaluate the general plethysm (see [60] for a review of symmetric functions and their various manipulations), in practice all required operations of symmetric functions involving products, plethysms and group branching rules can be evaluated symbolically using an appropriate group theory package. Where required, we use `Schur` [91] for this purpose.

From (8), which gives the form of the phylogenetic tensor for a tree with m leaves, it is clear that the appropriate representation space to consider is indeed $\otimes^m\mathbb{C}^k$, regarded not as a module of $GL(k)$ as above, but rather carrying an irreducible representation of the action of the direct product group $\times^m GL(k) = GL(k)\times GL(k)\times\ldots\times GL(k)$. That is, considering that a phylogenetic tensor lies in the ambient space $\otimes^m\mathbb{C}^k$, the generic analogue of (8) is

$$\psi' = (g_1\otimes g_2\otimes\ldots\otimes g_m)\cdot\psi, \qquad (12)$$

where $\psi\in\otimes^m\mathbb{C}^k$. In a phylogenetic setting, we must allow for differing Markov operators to act on each edge; hence the above form. It is usual in phylogenetics to take a fixed rate matrix for all edges, and allow the edge lengths to vary, thus creating different Markov operators from identical generators. In fact, the above generalization of the group action allows for differing Markov *processes* on every edge of the phylogenetic tree[7].

A complete representation-theoretic analysis incorporating the tree structure of phylogenetic tensors is a topic for future research, and we defer such a theory. We concentrate on analysing the group action defined by (8), leading toward the derivation of Markov invariants while ignoring the underlying tree structure. In §4 we will introduce a post-hoc procedure which allows the tree structure to be incorporated. This will allow Markov invariants to be applied in a practical setting without the need for the complete theory.

[6] In the context of quantum physics, this corresponds exactly to the statistical properties for ensembles of bosonic and fermionic particles, respectively.

[7] Under a (somewhat biologically unsound) model in which the evolution along the pendant edges occurs with identical transition probabilities, the group becomes the diagonal $GL(k)$ subgroup of the m-fold direct product group, and the representation reduces accordingly, precisely as in the initial discussion above.

3.3 Representations of $\times^m GL(k)$

Here we derive the group branching rule which is required to identify the irreducible modules under the group action (12).

There is yet another product of symmetric functions; the *inner product*, defined as

$$\{\lambda\} \odot \{\rho\} = \sum_{\sigma \vdash n} \gamma^{\sigma}_{\lambda\rho}\{\sigma\},$$

where $|\lambda| = |\rho| = |\sigma| = n$, and the $\gamma^{\sigma}_{\lambda\rho}$ are the integer multiplicities of occurrences of the σ representation in the Kronecker product representation between λ and ρ of the symmetric group $\mathfrak{S}_n$ [55].

Consider the direct product group $GL(k) \times GL(\ell)$, with group action on $V_1 \otimes V_2$, where V_1 is k-dimensional and V_2 is ℓ-dimensional, defined by $v_1 \otimes v_2 \mapsto g_1 v_1 \otimes g_2 v_2$. If the eigenvalues of g_1, g_2 are $x_1, x_2, \ldots, x_k$ and $y_1, y_2, \ldots, y_\ell$ respectively, then the character of this representation is the product

$$\{1\}(x) \cdot \{1\}(y) = (x_1 + \ldots + x_k)(y_1 + \ldots + y_l) = \{1\}(xy),$$

with

$$(xy) = (x_1y_1, x_1y_2, \ldots, x_2y_1, \ldots, x_ky_\ell).$$

Generalizing this result, consider the natural embedding, $GL(k) \times GL(\ell) \subset GL(k\ell)$, and the $\{\lambda\}$ representation of $GL(k\ell)$ restricted to the direct product group: $\Psi \mapsto \pi_\lambda(g_1 \times g_2)\Psi$ with $\Psi \in (V_1 \otimes V_2)^\lambda$. The character of this representation has decomposition

$$\{\lambda\}(xy) = \sum_{\rho,\sigma \vdash |\lambda|} \gamma^{\lambda}_{\rho\sigma}\{\rho\}(x) \cdot \{\sigma\}(y); \tag{13}$$

for details see [50, 88]. Thus, we see that the inner product plays an essential role in decomposing representations of the direct product group $GL(k) \times GL(\ell)$ into tensor products of irreducible modules of $GL(k)$ with irreducible modules of $GL(\ell)$:

$$(V_1 \otimes V_2)^\lambda = \sum_{\rho,\sigma \vdash |\lambda|} \oplus \gamma^{\lambda}_{\rho\sigma} V_1^{\rho} \otimes V_2^{\sigma}.$$

Presently we will use this result to derive branching rules for the group action that is relevant to phylogenetics (8).

In its general setting, a (group) *branching rule* describes the decomposition of a representation of a group, G, when restricted to a subgroup, $H \subset G$ (written as $G \downarrow H$) [87, Chap. V, §18]. For the present purpose, we consider $\times^m GL(k)$ as a subgroup of $GL(k^m)$, and given the defining representation of $GL(k^m)$, the corresponding branching rule is

$$GL(k^m) \downarrow \times^m GL(k): \quad \{1\} \longrightarrow \{1\} \otimes \{1\} \otimes \ldots \otimes \{1\} = \otimes^m \{1\}.$$

On the left-side of the arrow, $\{1\}$ denotes the defining representation of $GL(k^m)$, whereas on the right-side, $\{1\}$ denotes the defining representation of $GL(k)$.

If we take the generic $\{\lambda\}$ representation of $GL(k^m)$, the appropriate branching rule is[8]

$$GL(k^m) \downarrow \times^m GL(k): \quad \{\lambda\} \longrightarrow \sum_{\sigma_1,\sigma_2,\ldots,\sigma_m \vdash |\lambda|}^{\{\sigma_1\}\odot\{\sigma_2\}\odot\ldots\odot\{\sigma_m\}\ni\{\lambda\}} \{\sigma_1\} \otimes \{\sigma_2\} \otimes \ldots \otimes \{\sigma_m\}. \tag{14}$$

This result can be confirmed using the identity (13).

[8] This is a special case of a more general embedding $\{1\} \to \{\lambda_1\} \otimes \{\lambda_2\} \otimes \ldots \otimes \{\lambda_m\}$, for which each $\{\sigma_i\}$ in the decomposition is replaced by the appropriate plethysm $\{\lambda_i\}\underline{\otimes}\{\sigma_i\}$. For a recent discussion of the calculus of plethysms see [20].

The branching rule (14) gives the decomposition of irreducible modules of the time evolution at the pendant edges of a tree, as induced by (8), but considered as the representation of $\times^m GL(k) \subset GL(k^m)$ defined by

$$\Psi' = \pi_\lambda(g_1 \times g_2 \times \ldots \times g_m) \cdot \Psi,$$

for $\Psi \in (\otimes^m \mathbb{C}^k)^\lambda$.

In the setting of phylogenetics, we show in §3.4 that specializing to $\{\lambda\} \equiv \{d\}$ gives the decomposition of (homogeneous degree d) polynomials of phylogenetic tensors. In a practical setting, this corresponds exactly to taking (polynomial) transformations of the observed data set of character pattern counts. That is, recalling that the expectation value of character pattern counts in a sequence alignment is governed by a joint distribution on a tree corresponding to a phylogenetic tensor P, the above branching rule tells us how arbitrary polynomial functions of the character pattern counts decompose into components which transform among themselves under the time evolution given in (8). In addition to what is presented here, this potentially has application to any analysis involving pattern counts in molecular sequence data (see §5 for further comments on this matter).

In §3.5, we will exploit the branching rule directly, defining the one-dimensional modules in the decomposition (14) as *invariants*, and give existence conditions for Markov invariants. We must first establish the isomorphism between homogeneous degree d polynomials on a vector space V, and the module $V^{\{d\}}$.

3.4 Symmetric plethysms and invariants

Associated with any representation V of a group G is the so-called coordinate ring $\mathcal{P}(V)$ of polynomials[9] over $\mathbb{C}$ in the components $v_1, v_2, \ldots, v_k$, corresponding to a given basis for V. For such polynomials, $f(v)$, there is a natural group action,

$$f(v) \to g \cdot f(v) := f(g^{-1}v).$$

There is an isomorphism between the ring $\mathcal{P}(V)$ and the symmetric tensor algebra[10] $\bigvee(V)$:

$$\mathcal{P}(V) \equiv \sum_{d=0}^{\infty} \mathcal{P}^d(V) \cong \bigvee(V) \equiv \sum_{d=0}^{\infty} \bigvee^d(V), \tag{15}$$

with $\bigvee^d(V) \cong V^{\{d\}}$ and $\mathcal{P}^d(V)$ denoting the homogeneous polynomials of degree d. This reflects that an arbitrary homogeneous polynomial of degree d in k indeterminates can be specified by an array of determinates $f_{i_1 i_2 \ldots i_d}$ which is symmetric under permutation of indices:

$$f(v) = \sum_{1 \leq i_1, i_2, \ldots, i_d \leq k} f_{i_1 i_2 \ldots i_d} v_{i_1} v_{i_2} \ldots v_{i_d}.$$

Our interest in the above construction lies in the *invariant ring*, $\mathcal{P}(V)^G$, of polynomials that are invariant up to a multiplicative factor under the action of G, or more generally, for any subgroup $H \trianglelefteq G$,

$$f(hv) = \det(h)^w f(v), \tag{16}$$

for all $h \in H$ and $v \in V$. For matrix groups the multiplicative factor is the determinant with w denoting the *weight* of the invariant. Using the isomorphism (15), the identification of a linear basis of such invariants of degree d reduces to the identification, in the reduction of the $V^{\{d\}}$, of the one-dimensional representations of H in the branching rule $G \downarrow H$.

[9] $\mathcal{P}(V)$ is the ring of polynomials in the basis elements, $\xi_1, \xi_2, \ldots, \xi_k$, of the dual space V^* so that $\mathcal{P}(V) \equiv \mathbb{C}[\xi_1, \xi_2, \ldots, \xi_k]$ with $\xi_i(\delta_j) = \delta_{ij}$ for all $1 \leq i, j \leq k$.

[10] See [28, Chap. 4] for a discussion of the symmetric tensor algebra.

In particular, the one-dimensional representations of $GL(k)$ occur as follows. Note that the dimension of a representation is equal to the trace of the representing matrix of the identity. For the irreducible module V^{λ} this is given by $s_{\lambda}(1,1,\ldots,1)$. Thus, for a one-dimensional module, the corresponding Schur function must be monomial and (considering the definition of the Schur functions using summations over semi-standard tableaux given in [60]) this occurs only for partitions of the form $\{r^k\}$ for any integer $r > 0$. Additionally, considering that one-dimensional representations act by simply multiplying by the character itself, and that

$$\begin{aligned} s_{\{r^k\}}(x) &= (x_1x_2\ldots x_k)^r, \\ &= \det(g)^r, \end{aligned} \tag{17}$$

we see that, for any $\psi \in V^{\{r^k\}}$, we have

$$\psi \mapsto \det(g)^r\psi,$$

under the $\{r^k\}$ representation of $GL(k)$. This should be compared directly to (16).

Taking $\mathfrak{M}(k) \subset GL_1(k)$, we can construct Markov invariants by identifying polynomials lying in the invariant ring for $GL_1(k)$. Clearly, any polynomial

$$f \in \mathcal{P}(\otimes^m V)^{\times^m GL_1(k)}$$

must also satisfy (9) and is hence a Markov invariant. Recalling the salient subgroup chain (10), affiliated to the Markov semigroup, the representation-theoretic task is to evaluate the relevant branching rules for specific cases. The required branching rules derive from (14), together with identification of the specific form of one-dimensional representations of the subgroup in question.

It should be noted that this procedure leaves open the possibility that there exist Markov invariants that do not occur in the invariant ring for $GL_1(k)$. We leave this as an open problem, but note that it is plausible that such a possibility could be excluded by continuity arguments.

3.5 Markov invariants, existence theorems

Presently, we use the facts we have collected above to establish existence conditions for polynomial invariants for the group actions of $GL(k), GL_1(k)$ and $GL_{1,1}(k)$.

Theorem 1: Polynomial invariants for phylogenetic models.
Linearly independent polynomial invariants at degree d of the groups:

i. $\times^m GL(k)$,

ii. $\times^m GL_1(k)$, and

iii. $\times^m GL_{1,1}(k)$,

are given by the one-dimensional modules of these groups occurring in the decomposition of the $GL(k^m)$ module $(\otimes^m V)^{\{d\}}$. In each case the one-dimensional modules correspond to m-fold products of Schur functions labelled by partitions of d:

i. $\{r^k\} \otimes \{r^k\} \otimes \ldots \otimes \{r^k\}$,

ii. $\{r_1{+}s_1, r_1^{k-1}\} \otimes \{r_2{+}s_2, r_2^{k-1}\} \otimes \ldots \otimes \{r_m{+}s_m, r_m^{k-1}\}$, and

iii. $\{r_1{+}s_1, r_1^{k-2}, t_1\} \otimes \{r_2{+}s_2, r_2^{k-2}, t_2\} \otimes \ldots \otimes \{r_m{+}s_m, r_m^{k-2}, t_m\}$, respectively,

with

$kr \equiv d$,

$kr_a + s_a \equiv d$, and

$$(k-1)r_b + t_b + s_b \equiv d,$$

for all $1 \leq a, b \leq m$ respectively.

Given the isomorphism (15) and the branching rule (14) with $\{\lambda\} \equiv \{d\}$, in each case the number of admissible partitions of the given forms $\{\sigma_1\} \otimes \{\sigma_1\} \otimes \ldots \otimes \{\sigma_m\}$ is the number of times the inner product $\{\sigma_1\} \odot \{\sigma_2\} \odot \ldots \odot \{\sigma_m\}$ of irreducible representations of the symmetric group $\mathfrak{S}_d$ contains the one-dimensional irreducible representation $\{d\}$. This is also the number of linearly independent polynomial invariants in each case.

Proof: Each case identifies representations of $\times^m GL(k)$ with character $\{\sigma_1\} \otimes \{\sigma_2\} \otimes \ldots \otimes \{\sigma_m\}$, each component of which is a partition that corresponds to a one-dimensional representation of the respective subgroup. The dimension of this representation is the product of the dimension of each of the representations labelled by $\{\sigma_a\}$. Therefore the representation is one-dimensional if and only if, for each $\{\sigma_a\}$, the corresponding representation is one-dimensional.

For case (i), $GL(k)$, as we showed in §3.4, the representation labelled by $\{r^k\}$ is one-dimensional, providing an invariant of weight $w \equiv r$. For case (ii), $GL_1(k)$, it is established in the appendix that the representation of $GL(k)$ labelled by $\{r_a + s_a, r_a^{k-1}\}$ contains a unique one-dimensional module under $GL_1(k)$. For case (iii), as will also be established in the appendix, $GL_{1,1}(k)$ is isomorphic to $GL(k-1)$ and the $GL(k)$ character $\{r_a + s_a, r_a^{k-2}, t_a\}$ contains under branching to $GL(k-1)$, a unique one-dimensional module with character $\{r_a^{k-1}\}$.

□

Note that case (ii) is a special instance of case (iii), with $t_a = 0$, and case (i) is a special instance of case (ii), with $s_a = 0$. This reflects the definition (16).

Recall the inclusion

$$\mathfrak{M}(k) \subset GL_1(k) \lhd GL(k).$$

It is clear that any invariant that exists for case (i), with $w \equiv r$, or (ii), with $w \equiv r_1 = r_2 = \ldots = r_m$, is necessarily a Markov invariant, (9), with the particular form

$$f(P) = (\det M_1 \det M_2 \ldots \det M_m)^w f(\widetilde{P}). \tag{18}$$

In §4 we will count occurrences of this type of Markov invariant for various cases of interest to phylogenetics; $k = 2$ to 4 character states and trees with $m = 2$ to 10 leaves. We will also briefly review the algebraic structure of these invariants in the cases $m = 2$ to 4 when evaluated upon phylogenetic tensors, and give examples of how this structure can be gainfully employed in the problem of phylogenetic tree inference from molecular sequence data.

Taking case (ii) in its general form, we see that for $w_1 \equiv r_1, w_2 \equiv r_2, \ldots w_m \equiv r_m$, it is possible that there exist Markov invariants, taking the general form

$$f(P) = (\det M_1^{w_1} \det M_2^{w_2} \ldots \det M_m^{w_m})\, f(\widetilde{P}).$$

When the distinction is required, we refer to these invariants as *mixed weight Markov invariants.* In §4.3 we will show that such invariants do indeed exist in various cases of interest to phylogenetics. However, as yet the explicit form of these invariants has not been constructed, and their structure remains unexplored.

Recall the inclusion

$$\mathfrak{M}^*(k) \subset GL_{1,1}(k) \subset GL(k),$$

for the doubly-stochastic Markov semigroup. The case (iii) establishes existence conditions for polynomial invariants for this semigroup. These invariants will be valid for any joint distribution on a phylogenetic tree which is constructed using only doubly-stochastic matrices. This includes oft-used models such as Jukes-Cantor, K80, K3ST and SYM [94]. We report the above theorem, but defer the exploration of the invariants in this case.

4 Markov invariants in phylogenetics

In §4.1 we establish existence of Markov invariants relevant to phylogenetics for the cases of $k=2$ to 4 character states, distinguishing between true Markov invariants and invariants which are valid for the full general linear group. In §4.2 we report upon known algebraic relations between Markov invariants when evaluated upon phylogenetic tensors for $k=2$ to 4 character states and for trees with $m=2$ to 4 leaves. We also discuss the application of Markov invariants to the problem of phylogenetic tree reconstruction in these cases. Finally, in §4.3 we establish existence of mixed weight Markov invariants for $k=4$ character states and trees with $m=2$ to 5 leaves. Throughout, we used `Schur` [91] for all non-trivial manipulations of Schur functions.

4.1 Zoo of invariants and nomenclature

We gave, in §3.4, a sufficient condition for the existence of a Markov invariant, (18), of degree d and weight w:

$$\{r+s, r^{k-1}\} \odot \{r+s, r^{k-1}\} \odot \ldots \odot \{r+s, r^{k-1}\} \ni \{d\}, \tag{19}$$

where $r=w$ and the inner product is taken m times, subsequently written as $\odot^m\{r+s, r^{k-1}\}$.

For reasons discussed below, taking $r=0$ results in the trivial inner product:

$$\{s\} \odot \{s\} = \{s\},$$

for all integers $s > 0$. Extending to $m > 2$,

$$\odot^m\{s\} = \{s\},$$

and the corresponding Markov invariant is denoted as Φ with degree $d=1$ and weight $w=0$, and simply expresses the conservation of total probability under the action of the Markov semigroup:

$$\Phi(P) \equiv \sum_{i_1,i_2,\ldots,i_m} P_{i_1 i_2 \ldots i_m} = 1.$$

Here Φ is the invariant corresponding to $s=1$ and for $s>1$ the invariant is simply the power Φ^s.

For fixed m, and any two invariants f, f' of degree d, d' and weight w, w', we can form the pointwise product $f \cdot f'$ which is itself an invariant of degree $d+d'$ and weight $w+w'$. If $w=w'$, we can form an invariant from the sum $f+f'$. These statements establish that the invariants, $\mathcal{P}(V)^G$, form a *graded ring* [47] (where the grading is over both the degree d and the weights w). In particular, it is important to note that we can increase the degree of any invariant (keeping the weight fixed) by multiplying it with the trivial invariant Φ.

When searching for Markov invariants, we must note that the sufficiency condition (19) will include these powers, and hence in what follows we must allow for this over-counting. In the conclusions we will expand upon this observation with some comments in regard to classifying the ring of invariants.

The general linear, or $s=0$, case

Recalling Theorem 1, we see that for $s=0$, the Markov invariants are simultaneously invariants under the action of the general linear group. Taking $r=1$, the inner multiplication is trivial:

$$\{1^k\} \odot \{1^k\} = \{k\}.$$

This reflects that the Kronecker product of the *alternating* representation of $\mathfrak{S}_k$, associated with the partition (1^k), taken with itself, is the trivial representation, which in turn is associated with the partition (k). Recall that the alternating representation is one-dimensional whose action on $\mathbb{C}$ defined as multiplication by $+1$ if σ is an even permutation and is -1 otherwise. For this

	$k=2$		$k=3$		$k=4$	
m	$\{21\}$	$\{31\}$	$\{21^2\}$	$\{31^2\}$	$\{21^3\}$	$\{31^3\}$
2	1	1	1	1	1	1
3	1	1	1	1	0	1
4	3	4	4	13	4	16
5	5	10	10	61	6	137
6	11	31	31	397	40	1396
7	21	91	91	2317	126	13881
8	43	274	274	14029	568	138916
9	85	820	820	83917	2142	1388857
10	171	2461	2461	504013	8824	13888996

Table 1: Occurrences of $\{d\}$ in $\odot^m\{r+s, r^{k-1}\}$ with $rk+s=d$

one-dimensional representation, the Kronecker product is simply the numeric product, with the result being the trivial representation where every permutation is mapped to +1. Similarly

$$\{1^k\} \odot \{k\} = \{1^k\},$$

and we see that there exists a single Markov invariant of degree $d=k$ and weight $w=1$ for all *even* values of m.

A very familiar example occurs for $m=2$ where, as we will discuss in §4.2, the invariant arises as the Log-Det distance function [76]. In the next case, $m=4$, we refer to the corresponding Markov invariant as the *quangle*.

Considering $m=2$ and $r=2$, we have

$$\{2^k\} \odot \{2^k\} \ni \{2k\},$$

for $2 \leq k \leq 4$. For each k, these invariants can be accounted for by taking the previous invariant and multiplying by Φ. Thus nothing new is gained.

However, taking $m=3$, it follows that there exists an invariant of degree $d=2k$ and weight $w=2$:

$$\{2^k\} \odot \{2^k\} \odot \{2^k\} \ni \{2k\}.$$

For $k=2$ this invariant is known in the quantum physics literature as the *tangle* [15, 16], where it is drawn upon to classify entanglement in 3-qubit systems, and has been generalized for $k=3$ and 4 in the context of phylogenetics in [84]. In §4.2 we will briefly review the most striking properties of the tangle relevant to phylogenetics.

Bona-fide Markov invariants, $s>0$

Here we consider the case $s>0$, where the resulting Markov invariants are *not* simultaneously valid for the general linear group. In Table 1 we present the number of weight $w=1$ invariants that exist for the cases $k=2,3,4$; $m=2,3,\ldots,10$ and $s=1,2$. All required computations were performed using `Schur`, and we have not reduced for over-counting. In Table 2 we summarize the Markov invariants for which we have successfully computed explicit polynomial forms. Here we also record the nomenclature we have developed. Presently we discuss the particular properties of these invariants when evaluated on phylogenetic tensors derived from a tree.

4.2 What happens on a phylogenetic tree?

By definition, the expectation value of a (bias corrected) Markov invariant, f, depends only upon the internal part of the phylogenetic tree:

$$E[\widehat{f}(Z)] = f(P) = (\det M_1 \det M_2 \ldots \det M_m)^w f(\widetilde{P}),$$

Name	Symbol	Inner multiplication	Group	(d,w)
det	Det	$\odot^2\{1^2\} = \{2\}$	$\times^2 GL(2)$	(2,1)
		$\odot^2\{1^3\} = \{3\}$	$\times^2 GL(3)$	(3,1)
		$\odot^2\{1^4\} = \{4\}$	$\times^2 GL(4)$	(4,1)
tangle	T	$\odot^3\{2^2\} \ni \{4\}$	$\times^3 GL(2)$	(4,2)
		$\odot^3\{2^3\} \ni \{6\}$	$\times^3 GL(3)$	(6,2)
		$\odot^3\{2^4\} \ni \{8\}$	$\times^3 GL(4)$	(8,2)
stangle	T^s	$\odot^3\{21\} \ni \{3\}$	$\times^3 GL_1(2)$	(3,1)
		$\odot^3\{21^2\} \ni \{4\}$	$\times^3 GL_1(3)$	(4,1)
		$\odot^3\{31^3\} \ni \{6\}$	$\times^3 GL_1(4)$	(6,1)
quangle	Q	$\odot^4\{1^2\} \ni \{2\}$	$\times^4 GL(2)$	(2,1)
		$\odot^4\{1^3\} \ni \{3\}$	$\times^4 GL(3)$	(3,1)
		$\odot^4\{1^4\} \ni \{4\}$	$\times^4 GL(4)$	(4,1)
squangle	Q^s	$\odot^4\{21\} \ni 3\{3\}$	$\times^4 GL_1(2)$	(3,1)
		$\odot^4\{21^2\} \ni 4\{4\}$	$\times^4 GL_1(3)$	(4,1)
		$\odot^4\{21^3\} \ni 4\{5\}$	$\times^4 GL_1(4)$	(5,1)

Table 2: Markov invariants of degree d and weight w for m leaves,

where Z is the observed counts of character patterns, P is the phylogenetic tensor corresponding to the joint distribution on the tree, and the trimmed tensor $\widetilde{P}$, defined in §2.4, is formed by setting the lengths of the pendant edges to zero. It is exactly this property that can be exploited in the practical setting of reconstructing phylogenetic trees from molecular sequence data.

As discussed in the closing comments of §3.2, Markov invariants exist independently of any notion of a tree, and to uncover their potential use in the problem of phylogenetic tree reconstruction it becomes necessary to analyse their structure on particular trees. Crucial to this examination is the *generalized pulley principle* presented in [84], which establishes that the family of probability distribution resulting from taking the general Markov model on a particular tree is unchanged under arbitrary placement of the root of the tree (see [1] for an equivalent discussion). Thus, our task is to search for algebraic relations between the Markov invariants valid for a given m, when evaluated upon the trimmed phylogenetic tensors corresponding to particular trees with m leaves. We are free to place the root arbitrarily, and we choose to evaluate the Markov invariants on trees where the root is located to our convenience.

In Appendix B we present the general procedure for computing the explicit polynomial form of Markov invariants using the Young's operators (§3.2) associated with the relevant partitions. Our general procedure was to take these explicit forms and then search for algebraic relations when the invariants are evaluated on the pruned tensor $\widetilde{P}$ defined by a particular tree. In the general case, any such relations potentially lead to phylogenetically informative statistics, valid under a general model of sequence evolution. Presently we will report upon this procedure in the known cases, $m = 2, 3$ and 4.

The simplest Markov invariant: the Log-Det

Recall that the generic phylogenetic tensor on $m\!=\!2$ leaves (Figure 1) can be written in the form

$$P = (M_1 \otimes M_2) \cdot (\delta \cdot \pi).$$

The corresponding trimmed tensor, $\widetilde{P}\!=\!\delta \cdot \pi$, can be expressed in the $\delta_1, \delta_2, \ldots, \delta_k$ basis with the components

$$\widetilde{P}_{i_1 i_2} = \delta_{i_1 i_2} \pi_{i_1}.$$

As we showed above, there exists a single Markov invariant for $m\!=\!2$. The polynomial form of this invariant is easily derived by considering rank 2 tensors as matrices, and taking the determinant.

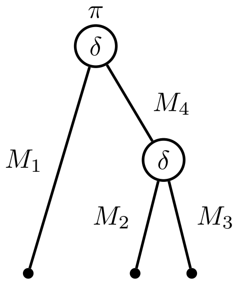

Figure 6: Phylogenetic tensor for the tree (1,23)

Since the invariant is a function on tensors, we make the distinction by using a capital letter and denoting the invariant as Det. This distinction can be compared directly to the use of the determinant function in [4] as opposed to the use in [76].

Substitution gives

$$\mathrm{Det}(\widetilde{P}) = \prod_{1\leq i\leq k} \pi_i,$$

such that, by the definition of Det as a Markov invariant,

$$\mathrm{Det}(P) = \det(M_1)\det(M_2) \prod_{1\leq i\leq k} \pi_i. \tag{20}$$

This form holds for *any* k, and is exploited by taking the logarithm and computing the Log-Det distance measure [52, 59].

Triplet distances: the tangle

Inspection of Table 1 reveals that for $m=3$ and $s=0$ there exists a Markov invariant, for each of $k=2$, 3 and 4, of degree $d=2k$ and weight $w=2$. This invariant is valid for phylogenetic trees with three leaves. For each of $k=2$, 3 and 4, the explicit polynomial forms of the tangle are basis independent (by definition) and have 12, 1152 and 431424 terms respectively.

The generic phylogenetic tensor on the three leaf tree (Figure 6) can be expressed as

$$P = (1\otimes M_2\otimes M_3)\cdot(1\otimes\delta)\cdot(M_1\otimes M_4)\cdot(\delta\cdot\pi). \tag{21}$$

The trimmed tensor, $\widetilde{P} = (1\otimes\delta)\cdot(1\otimes M_4)\cdot(\delta\cdot\pi)$, has components

$$\widetilde{P}_{i_1i_2i_3} = P^*_{i_1i_2}\delta_{i_2i_3}, \tag{22}$$

where $P^*=(1\otimes M_4)\cdot(\delta\cdot\pi)$ is the pruned tensor.

The tangle is a Markov invariant and hence satisfies

$$T(P) = (\det M_1 \det M_2 \det M_3)^2 T(\widetilde{P}).$$

By explicit computation we have found that, for each of $k=2$, 3 and 4,

$$T(\widetilde{P}) = \mathrm{Det}^2(P^*).$$

Thus we see that the induced function of the tangle is $T^*\equiv \mathrm{Det}^2$. This is the example we promised in §2.5.

Consistent with (20) we have

$$\mathrm{Det}(P^*) = \det M_4 \left(\prod_{1\leq i\leq k}\pi_i\right),$$

so, finally, we see that

$$T(P) = (\det M_1 \det M_2 \det M_3 \det M_4)^2 \left(\prod_{1 \le i \le k} \pi_i \right)^2 . \qquad (23)$$

Due to the generalized pulley principle, (23) holds for the phylogenetic tensor corresponding to *any* tree with three leaves. Comparing directly to (20) it is clear that the tangle may be used similarly to the Log-Det pairwise distance but for triplets of molecular sequence data. For further details in this direction see [84].

Informative statistic: the stangle

We see from Table 1 that for $m=3$ and $s=2$ there also exists, for each $k = 2$, 3 and 4, a weight $w=1$ Markov invariant valid for trees with three leaves (of degree $d=6$ for $k=4$ states). We refer to this invariant as the *stangle*, that is, the *s*tochastic *tangle* (see [81] for explicit expressions for the $k=2$ and 3 cases). As discussed in Appendix B, the explicit polynomial form of the stangle for $k=4$ is known only in a basis different from the standard $\delta_1, \delta_2, \ldots, \delta_k$. In this basis, the stangle has 1404 terms with relevant data files available on Charleston's website [82]. This does not, however, prevent us from using the stangle in a practical setting as evaluation can be performed in this basis by transforming the *data set* (pattern counts) into the required basis.

For the trimmed tensor with components given by (22), explicit computation shows that the stangle satisfies $T^s(\widetilde{P}) \equiv 0$. Thus the stangle is simultaneously a *phylogenetic invariant* for a tree with three leaves (of course this again holds for *any* tree with three leaves).

Given an unbiased estimator, $\widehat{T}^s$, of the stangle, we see that under the family of probability distributions described by (21), the expectation value of this estimator when evaluated on triplets of aligned DNA sequences is zero :

$$E[\widehat{T}^s(Z)] = 0,$$

where Z is the tensor of observed pattern counts in the aligned sequence data. Deviation from zero by the observed value of the stangle can thus be viewed as evidence that the data set violates the assumptions of the Markov model. We have had some preliminary (unpublished) success capitalizing on this property to rank subsets of aligned molecular sequences according to apparent concurrence with model assumptions.

Note that the stangle must occur within the framework of phylogenetic invariants presented in [1] and the discussion of [53]. It would be interesting to determine whether the stangle is a linear combination (with coefficients that are $d=1$ polynomials) of the quintic phylogenetic invariants presented in [79]. However, whether or not this is the case is beyond the theoretical techniques presented in this paper and more work needs to be done before the precise connections between the stangle and the known phylogenetic invariants for this case become transparent. Further, because the explicit polynomial form of the stangle in the standard basis is not known, brute-force determination is impractical using algorithms presently known to the authors.

Quartet inference: the squangles

Inspection of Table 1 reveals that for $k=4$ and $m=4$, there exist four Markov invariants of degree $d=5$ and weight $w=1$ relevant to phylogenetic trees with four leaves. We refer to these invariants as the *squangles*. Again, the explicit polynomial form of the squangles is known only in a basis different from the standard one, and data files can be found on Charleston's website [82]. We have found that three particular linear combinations of the squangles are tree informative. Here we denote these three squangles as Q_1, Q_2 and Q_3. In the non-standard basis, Q_1 has 77004 terms, whereas both Q_2 and Q_3 have 91620 terms.

On the quartet tree in Figure 7, the generic phylogenetic tensor is

$$P = (M_1 \otimes M_2 \otimes M_3 \otimes M_4) \cdot (\delta \otimes \delta) \cdot (M_5 \otimes M_6) \cdot (\delta \cdot \pi).$$

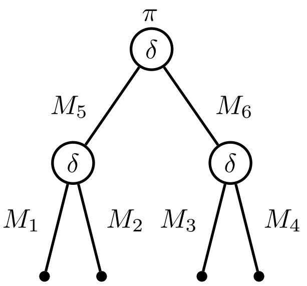

Figure 7: Phylogenetic tensor for the tree (12,34)

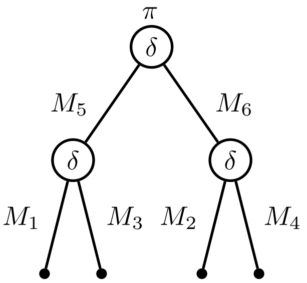

Figure 8: Phylogenetic tensor for the tree (13,24)

The trimmed tensor $\widetilde{P} = (\delta \otimes \delta) \cdot (M_5 \otimes M_6) \cdot (\delta \cdot \pi)$ has components:

$$\widetilde{P}_{i_1 i_2 i_3 i_4} = P^*_{i_1 i_3} \delta_{i_1 i_2} \delta_{i_3 i_4},$$

with the pruned tensor given by $P^* = (M_5 \otimes M_6) \cdot (\delta \cdot \pi)$. This form of the trimmed tensor can be evaluated directly on the explicit polynomial form of the squangles. We found that on the tree $(12, 34)$ the squangles satisfy the algebraic relations:

$$Q_1(\widetilde{P}) = 0, \quad Q_2(\widetilde{P}) = -Q_3(\widetilde{P}) > 0,$$

with, intriguingly, the polynomial form of $Q_2(\widetilde{P})$ with respect to the components $P^*_{i_1 i_2}$ taking that of the *permanent*[11], which, unfortunately for the phylogenetic context, is not a Markov invariant.

An identical procedure was carried out on the phylogenetic tensors corresponding to the trees in Figure 8 and Figure 9. This produced the relations

$$\begin{aligned} Q_2(\widetilde{P}) = 0, &\qquad Q_1(\widetilde{P}) = Q_3(\widetilde{P}) > 0, \\ Q_3(\widetilde{P}) = 0, &\quad -Q_1(\widetilde{P}) = -Q_2(\widetilde{P}) > 0, \end{aligned}$$

respectively.

Noting these relations, we see that we have constructed tree-informative phylogenetic invariants for trees with four leaves. In particular, for the unbiased estimators thereof, we have

$$E[\widehat{Q}_1(Z)] \equiv 0, \qquad E[\widehat{Q}_2(Z) + \widehat{Q}_3(Z)] \equiv 0;$$

for the tree $(12, 34)$,

$$E[\widehat{Q}_2(Z)] \equiv 0, \qquad E[\widehat{Q}_1(Z) - \widehat{Q}_3(Z)] \equiv 0;$$

for the tree $(13, 24)$, and

$$E[\widehat{Q}_3(Z)] \equiv 0, \qquad E[\widehat{Q}_1(Z) - \widehat{Q}_2(Z)] \equiv 0;$$

[11] The permanent has identical algebraic form to the determinant of a matrix but with each term replaced by its absolute value.

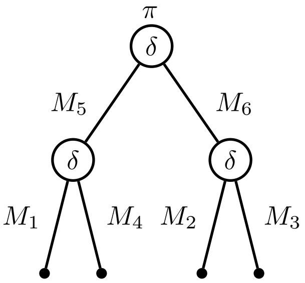

Figure 9: Phylogenetic tensor for the tree (14,23)

for the tree $(14, 34)$. We also note that the linear combination

$$W := Q_1 - Q_2 - Q_3,$$

satisfies $E[\widehat{W}(Z)] \equiv 0$, for *any* phylogenetic tree with four leaves.

The bar charts in Figure 10 compare the success of three tree inference methods tested on data sets created using `Hetero` [43]. All parameter settings used are as presented in [44] with sequence length N=10000 and 10000 runs being completed in each case. A molecular clock was imposed, and for each run the tree used to simulate the data was

$$\texttt{tree1} = ((Seq1 : 0.495, Seq2 : 0.495) : 0.005, (Seq3 : 0.495, Seq4 : 0.495) : 0.005),$$

with branch lengths given in time units and t=0.495 and .005 corresponding to 0.1485 and 0.0015 expected number of state changes, respectively. The G+C content was made to increase in leaves 1 and 4 and was reduced in leaves 2 and 3. This tends to bias tree inference to `tree3`= $(14, 23)$ as sequences 1 and 4 will tend to be more similar purely because of the G+C content.

The Maximum Likelihood and Log-Det+NJ quartet inferences were performed using the default settings in `Phylip` [23], whereas the Log-Det+BIONJ inferences were performed using the `R` [71] package "`ape`" [68]. Finally, the squangles inferences were implemented in `R` using our own original code [82]. For the purpose of making a rough comparison, on average each evaluation took .58s for maximum likelihood, .036s for Log-Det+NJ, .090s for Log-Det+BIONJ, and .085s for the squangles. The squangles routine was designed for illustrative purposes only and was performed under simple statistical assumptions, as follows.

The squangles were taken to be stochastically independent and normally distributed, with identical variances, σ^2, and mean values set to 0 or $u > 0$, depending on the quartet under consideration and the expectation values given above. That is, for each quartet in turn, we took

$$\begin{aligned}
\mathbb{P}[Q_1, Q_2, Q_3|(12, 34)] &\sim \mathcal{N}(0, \sigma^2) * \mathcal{N}(u, \sigma^2) * \mathcal{N}(-u, \sigma^2),\\
\mathbb{P}[Q_1, Q_2, Q_3|(13, 24)] &\sim \mathcal{N}(u, \sigma^2) * \mathcal{N}(0, \sigma^2) * \mathcal{N}(u, \sigma^2),\\
\mathbb{P}[Q_1, Q_2, Q_3|(14, 23)] &\sim \mathcal{N}(-u, \sigma^2) * \mathcal{N}(-u, \sigma^2) * \mathcal{N}(0, \sigma^2).
\end{aligned}$$

Our primary scientific justification for these assumptions is that the resulting quartet inference routine performs rather well.

Under these assumptions the maximum likelihood estimate (MLE) of u is independent of σ^2, and is equivalent to the least squares estimator. *Analytic* solutions are easily derived:

$$\begin{aligned}
\text{MLE}\,[u|(12, 34)] &= \max\left[0, \frac{Q_2(Z) - Q_3(Z)}{2}\right],\\
\text{MLE}\,[u|(13, 24)] &= \max\left[0, \frac{Q_1(Z) + Q_3(Z)}{2}\right],\\
\text{MLE}\,[u|(14, 23)] &= \max\left[0, \frac{-(Q_1(Z) + Q_2(Z))}{2}\right].
\end{aligned}$$

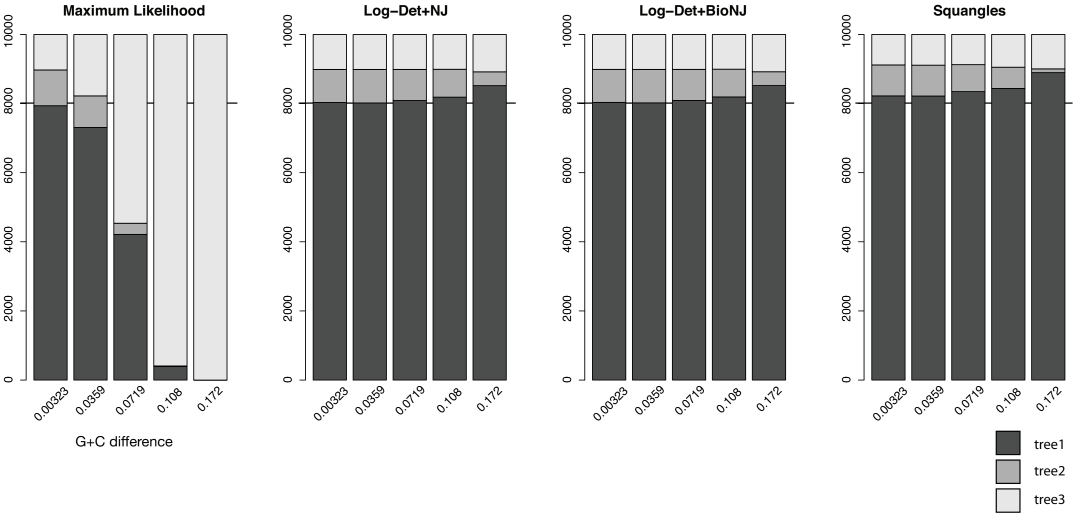

Figure 10: **Quartet reconstruction using the squangles.** The charts present how many times the `tree1`=(12,34), `tree2`=(13,24) and `tree3`=(14,23) were reconstructed using each of the three methods displayed. The tree used to simulate the data was `tree1`.

For each data set and candidate quartet, we computed the MLE for the mean value u and chose the quartet with the maximum likelihood.

While our demonstration is not intended as an exhaustive comparison between the performance of our method and ML using the default settings of `Phylip`, it does show that using a stationary model for ML can lead to incorrect tree inference if the data was produced by a non-stationary process. With that caveat, it is clear that ML performs very badly as the G+C content increases, strongly favouring `tree3`. The Log-Det routine is robust against varying G+C content as the technique is based on a general model (this is consistent with what was found in [44]). Being valid for a general model, the squangles are also robust against varying G+C content, and actually perform slightly better than Log-Det.

Interestingly, as the G+C content increases, the Log-Det and the squangles infer the true tree *more* often. Careful inspection reveals that this is because, as the G+C content increases, both these techniques tend to infer `tree2` less often and `tree3` at approximately the same rate, favouring `tree1`. This effect is more pronounced for the squangles.

4.3 Mixed weight Markov invariants

Here we report upon the existence of some mixed weight invariants for various cases of interest to phylogenetics. The polynomial form and algebraic structure on trees of these invariants remains completely unexplored.

We concentrate on $k=4$ and look for mixed weight invariants for the degree $d=8$ partition shapes $\{2^4\}$ and $\{51^3\}$, corresponding to $s=0$ and 4 respectively.

In the $m=2$ case, we find that

$$\{2^4\}\odot\{51^3\}$$

does not contain $\{8\}$, which means there does not exist a mixed weight invariant for trees on two leaves.

In the $m=3$ case, we have

$$\{2^4\}\odot\{51^3\}\odot\{51^3\}\ni\{8\},$$
$$\{2^4\}\odot\{2^4\}\odot\{51^3\}\ni\{8\}.$$

Writing $w=(w_1,w_2,w_3)$, we see that, including the three possible permutations across the inner products, there exist mixed weight invariants for the cases $w=(2,1,1),(1,2,1),(1,1,2)$ and $w=(2,2,1),(2,1,2),(1,2,2)$ respectively.

In the $m=4$ case, we have

$$\begin{aligned}\{2^4\}\odot\{51^3\}\odot\{51^3\}\odot\{51^3\}&\ni 14\{8\},\\ \{2^4\}\odot\{2^4\}\odot\{51^3\}\odot\{51^3\}&\ni 9\{8\},\\ \{2^4\}\odot\{2^4\}\odot\{2^4\}\odot\{51^3\}&\ni 4\{8\}.\end{aligned}$$

Taking account of the permutations, we see that there exist $14\times 4=54$, $9\times 6=54$ and $4\times 4=16$ mixed weight invariants for the cases $w=(2,1,1,1)$, $w=(2,2,1,1)$ and $w=(2,2,2,1)$ respectively.

Finally, in the $m=5$ case, we have

$$\begin{aligned}\{2^4\}\odot\{51^3\}\odot\{51^3\}\odot\{51^3\}\odot\{51^3\}&\ni 527\{8\},\\ \{2^4\}\odot\{2^4\}\odot\{51^3\}\odot\{51^3\}\odot\{51^3\}&\ni 212\{8\},\\ \{2^4\}\odot\{2^4\}\odot\{2^4\}\odot\{51^3\}\odot\{51^3\}&\ni 90\{8\},\\ \{2^4\}\odot\{2^4\}\odot\{2^4\}\odot\{2^4\}\odot\{51^3\}&\ni 46\{8\}.\end{aligned}$$

Again taking account of the permutation, we see that there exist $527\times 5=2635$, $212\times 10=2120$, $90\times 10=900$ and $46\times 5=230$ mixed weight invariants for the cases $w=(2,1,1,1,1)$, $w=(2,2,1,1,1)$, $w=(2,2,2,1,1)$ and $w=(2,2,2,2,1)$ respectively.

We expect that a future analysis of the explicit form of these invariants will lead to quite an array of informative statistics for phylogenetics.

5 Discussion

In this work we have defined and explored the construction of 'Markov invariants'. The primary tool exercised was group representation theory, applied to the usual Markov process present in probabilistic models of phylogenetic trees.

It is evident that our present approach to phylogenetics offers many possibilities for further study. The various Markov invariants that we have identified and constructed provide strong candidates for improved tree estimation and parameter recovery under general model assumptions. In particular, the stangle (§4.2) seems to provide a robust indicator of phylogenetic signal in subsets of aligned molecular sequences. Efforts are underway to incorporate the stangle into a clustering algorithm that provides the means to divide large phylogenetic data sets into smaller, manageable parts, inspired, in part, by the Disk-Covering technique of [37]. In §4.2, we presented a Markov invariant based quartet inference technique. The maximum likelihood estimation we employed was based on rather simple statistical assumptions, and it is clear that this technique could easily be improved upon. Detailed knowledge of the invariants' distribution is desirable not only in order to achieve correct tree inferences, but also to find confidence intervals (as [61] do for Log-Det and ML distances). In all its glorious detail, the joint distribution of the squangles can be derived using the multinomial distribution of Z:

$$\mathbb{P}[Q_1(Z)=q_1,Q_2(Z)=q_2,Q_3(Z)=q_3]=\sum_{z\in\Upsilon}\mathbb{P}[Z=z;N]=N!\sum_{z\in\Upsilon}\left(\prod_{I\in K^m}\frac{\mu_I^{z_I}}{z_I!}\right),$$

where the summation is over the variety $\Upsilon:=\{z|Q_1(z)-q_1=Q_2(z)-q_2=Q_3(z)-q_3=0\}$. However, this distribution depends implicitly upon the model parameters underlying μ, and therefore negates the whole point of employing invariants in the first place! Clearly, a more coarse grained approach is desirable for intuiting an approximate distribution for the invariants that depends on just a few shape parameters. This can be achieved variously by studying the relevance and impact of the central limit theorem, deriving the first few moments using the generating function (2), or conducting extensive simulation studies. This would help to provide rigorous justification for taking the invariants as normally distributed, as we did for the squangles in §4.2.

Citing poor performance on short sequences [33, 35], there is a somewhat popular opinion that phylogenetic invariants are of limited utility when it comes to phylogenetic inference in practice. However, recent work suggests that this performance can be greatly improved by identifying "powerful" invariants [17]. For instance, [12] chose invariants for the K3ST model using a criterion arising from algebraic geometry, and [18] used a learning algorithm to choose invariants for the K3ST and Jukes-Cantor models. However, determining criteria that guarantee identification of statistically powerful invariants is in general an outstanding problem. In this context, we have shown clearly that Markov invariants can be of significant practical utility. For instance, one need only note that the simplest Markov invariant forms the structure of the Log-Det distance measure, an extremely popular tool employed in countless phylogenetic studies, while the simulation study we presented in §4.2 shows that Markov invariants can be used to infer quartet phylogenies with a success rate equivalent to, or greater than, popular methods. For phylogenetic invariants that arise as Markov invariants, it would be interesting to determine whether the additional analytic structure imposed by group invariance provides an effective criterion for identification of powerful invariants.

Markov invariants occur as one-dimensional representations of a group action associated with the Markov semigroup. In this regard, we applied only a particular instance of the group branching rule (14) that requires each of the irreducible modules to be one-dimensional. The standard approach to maximum likelihood exploits the trivial instance of the same branching rule with $\{\lambda\}=\{\sigma_1\}=\{\sigma_2\}=\ldots=\{\sigma_m\}\equiv\{1\}$, taking m copies of the k-dimensional *defining* representation to obtain a k^m-dimensional and degree $d=1$ polynomial representation. From this perspective, the standard approach and the Markov invariants are simply two cases where the transformation properties of polynomials of molecular sequence data under the time evolution (8) are exploited. This begs the question whether there exist polynomial representations, of dimension other than these two extremes, that can also be effectively utilized in practical phylogenetic tree inference.

Many of the different classes of phylogenetic models [36] can be affiliated with appropriate subgroups of $GL(k)$, and can therefore be expected to have a place in the subgroup chain (10). In principle we can modify Theorem 1 (§3.5) for each of these models and construct their associated Markov-type invariants. In this vein, Appendix C outlines a group-theoretic analysis of the Kimura 3ST model connecting the Hadamard conjugation with the construction of the Cartan subalgebra for this model. The same considerations apply in principle to amino acid sequence models: this is simply a matter of setting $k=20$ and using the same theory, though the computations involved will of course be more lengthy.

As noted in §3.2, a representation-theoretic analysis of the space of phylogenetic tensors that includes the underlying tree structure has not been developed in this work. Ideally, for a given tree, one would like to obtain the structure of the ring of Markov invariants as a theoretical outcome, rather than obtain this structure using the post-hoc procedure presented in §4.2. A possible direction in this regard is to consider, for each tree with m labelled leaves, the subgroup of $\mathfrak{S}_m$ induced by identifying permutations that leave the leaf labelling invariant. This subgroup is discussed in [72, Chap. 2] and, in a different context, in [6, Chap. 12, Topic 3]. We conjecture that this group may play a role, analogous to that of the symmetric group for the Schur-Weyl duality, in the construction of the irreducible modules for the space of phylogenetic tensors.

The phylogenetic invariants form an *ideal* in the associated polynomial ring, and hence Hilbert's basis theorem for finite-generatedness applies. However, whether Markov invariants are finitely-generated is unknown. Technically, the group $GL_1(k)$ is *non-reductive* (its finite-dimensional representations are not completely reducible). In the non-reductive case, standard theorems, such as finite-generatedness of polynomial invariant rings, do not apply. Thus it is unlikely that the Markov-type invariants will be finitely generated in general. A notable exception is provided by Weitzenböck's theorem [73, 86] for finite-dimensional representations of one-dimensional Lie groups. Continuing the subgroup chain (10) to its natural limit, it follows that in the case of phylogenetic tensors, the group $GL_1(k)$ provides, on restriction, an indecomposable representation of the additive group $\mathbb{R}^+$ (corresponding to time evolution). Thus, Weitzenböck's theorem is relevant to the analysis of Markov invariants in the current context. In fact, this observation is pertinent to phylogenetic invariants for continuous time models, as they would occur as *syzygies* [66, Chap. 2] between invariants belonging to the generating set of the invariant ring for the representation of $\mathbb{R}^+$ in question.

Acknowledgments
This paper is the culmination of several years work resulting from interactions with many researchers in what is, for Sumner and Jarvis, an unfamiliar field. There are many people to acknowledge for their various contributions ranging from simple positive encouragement to important technical insights. In this regard, we would like to thank Michael Baake, Peter Forrester, Alexei Drummond, David Bryant, Mike Steel, David Penny, Mike Hendy, Susan Holmes, Mark Pagel, Andreas Dress, Elizabeth Allman, John Rhodes, John Robinson, Bertfried Fauser, Ron King, Mike Eastwood, Jim Bashford, Malgorzata O'Reilly, and Simon Wotherspoon.

This research was conducted with support from the Australian Research Council grants DP0344996, DP0770991 and DP0877447.

A Proof of Theorem 1

We provide a tensor-based completion of the proof of Theorem 1 (§3.5), regarding the identification of one-dimensional irreducible representations of the groups $GL(k)$, $GL_1(k)$ and $GL_{1,1}(k)$.

Following the notation of §2, a probability measure can be written in a basis of point measures, $\mu = \sum_{1\leq i\leq k} \mu_i \delta_i$, with the Markov semigroup acting as

$$\mu \mapsto M\mu, \qquad M\delta_i = \sum_{1\leq j\leq k} \delta_j M_{ji}, \qquad M\mu = \sum_{1\leq i\leq k} {\mu'}_i \delta_i, \qquad {\mu'}_i = \sum_{1\leq j\leq k} M_{ij}\mu_j. \tag{A-1}$$

Moreover, probability conservation requires the column-sum condition $\sum_{1\leq i\leq k} M_{ij} = 1$, for all $1 \leq j \leq k$. As discussed in §3, this affiliates the linear transformations $M \in \mathfrak{M}(k)$ with the subgroup $GL_1(k) \lhd GL(k)$. Correspondingly, a higher rank tensor, ψ, transforms under the action of $g \in GL_1(k)$, $\psi \mapsto \psi'$, with

$${\psi'}_{i_1 i_2 i_3 \ldots} = \sum_{1\leq j_1, j_2, j_3 \ldots \leq k} g_{i_1 j_1} g_{i_2 j_2} g_{i_3 j_3} \ldots \psi_{j_1 j_2 j_3 \ldots}.$$

In order to find combinations of $\psi_{i_1 i_2 i_3 i_4 \ldots}$ which remain invariant up to scaling under the $GL_1(k)$ action, we transform to a more convenient basis in which the distinguished role of the vector $(1, e^\top) = (1, 1, \ldots 1)$ is identified. Following [65], define a nonsingular $k \times k$ matrix, X, with $1 \times 1 + (k-1) \times (k-1)$ block decomposition:

$$X := \begin{pmatrix} 1 & e^\top \\ \eta & x \end{pmatrix}. \tag{A-2}$$

Lemma: With respect to the similarity transformation $g \mapsto \widetilde{g} = XgX^{-1}$ defined by any fixed X of the above form, $GL_1(k)$ is isomorphic to the affine group $A(k) \cong GL(k-1) \ltimes T(k-1)$. Furthermore, under the same mapping subject to the constraint $\eta = -x \cdot e$, $GL_{1,1}(k)$ is isomorphic to the group $GL(k-1)$.

Proof: Check explicitly that

$$\text{if} \qquad g = \begin{pmatrix} \lambda & \ell_1^\top \\ \ell_2 & m \end{pmatrix}, \qquad \text{then} \qquad XgX^{-1} = \begin{pmatrix} 1 & 0 \\ \widetilde{\ell} & \widetilde{m} \end{pmatrix},$$

using the column-sum condition on g. Clearly, $\det g = \det \widetilde{m}$, so $\widetilde{m} \in GL(k-1)$ for all such X. Finally, if $\ell_2 = 1 - m \cdot e$, $\lambda = 1 - \ell_1^\top \cdot e$ and $\eta = -x \cdot e$, then $\widetilde{\ell} = 0$ in XgX^{-1} and $g \in GL_{1,1}(k)$ is thereby identified with the $GL(k-1)$ subgroup of $GL(k)$ consisting of matrices in block form as displayed. □

It is convenient to re-label the basis as $X\delta_1 := \widetilde{\delta}_0 = \delta_1 + \delta_2 + \ldots + \delta_k$, $X\delta_a := \widetilde{\delta}_a$, $\underset{\sim}{a} = 2, 3, \ldots, k$. In the new basis, probability measures will transform *in*homogeneously, with the $\widetilde{\delta}_0$ components invariant; for example mimicking (A-1)

$$\widetilde{\mu}'_0 = \widetilde{\mu}_0, \qquad \widetilde{\mu}'_a = \widetilde{\ell}_a\, \widetilde{\mu}_0 + \sum_{b=2}^{k} \widetilde{m}_{ab}\, \widetilde{\mu}_b\,, \tag{A-3}$$

and in this way we can deduce the transformation properties of higher-rank tensors.

As we discussed in §3.2, the finite dimensional *irreducible* representations of $GL(k)$ associated with partitions λ, are realized by tensors of rank $|\lambda|$ whose indices satisfy particular symmetrization conditions to be outlined in Appendix B: symmetrize across the rows and then anti-symmetrize down the columns of the associated standard tableau $\mathfrak{T}$. Conventionally, for example, we write for such a tensor the components $\psi_{[i_1 i_2 \ldots][j_1 j_2 \ldots][\ldots]}$. Here the indices enclosed in braces $[\ldots]$ are mutually anti-symmetric, corresponding to column entries in $\mathfrak{T}$, and there are further cyclic identities (we need not consider) reflecting row dependencies of ψ.

Below we will discuss properties of such tensors in the $\widetilde{\delta}_0, \widetilde{\delta}_2, \ldots \widetilde{\delta}_k$ basis under the transformation (A-3). The crucial result will depend absolutely on the indices, and the symbol 'ψ' will be superfluous. Hence, for ease of reading we will suppress the 'ψ':

$$\psi_{[i_1 i_2 \ldots][j_1 j_2 \ldots][\ldots]} \equiv [i_1 i_2 \ldots][j_1 j_2 \ldots][\ldots].$$

This is consistent with the amusing comments in the preface of [63].

Consider the reduction of an irreducible representation λ of $GL(k)$ with respect to the subgroup $GL(k-1)$ (equivalent, by the Lemma above, to considering the restriction to $GL_{1,1}(k)$ affiliated to the doubly-stochastic Markov semigroup). The partition labels $\overline{\lambda}$ of irreducible representations of $GL(k-1)$ arising from this restriction are related to those of λ by the standard betweenness conditions [87, Chap. V, §18] (see also [7, 88]):

$$\lambda_1 \geq \overline{\lambda}_1 \geq \ldots \geq \overline{\lambda}_{n-1} \geq \lambda_n. \tag{A-4}$$

Our present purpose is to identify one-dimensional representations of $GL(k–1)$, that may extend to one-dimensional representations of $GL_1(k) \cong GL(k–1) \ltimes T(k–1)$. Such tensor representations must be associated with partitions $\overline{\lambda} = (r^{k–1})$ all of whose columns have length $k-1$ corresponding to the r^{th} power of the representation $M \mapsto \det M$. However, for such a $\overline{\lambda}$, (A-4) above immediately implies that

$$\lambda = (r+s, r^{k-2}, t), \qquad \text{for some} \quad s \geq 0,\, t \leq r,$$

and we have established part (iii) of Theorem 1.

Within such tensor representations of type $(r+s, r^{k-2}, t)$, the component associated with the scalar representation $(r^{k–1})$ of $GL(k-1)$ is clearly

$$[0a_{12}\ldots a_{1k}][0a_{22}\ldots a_{2k}]\ldots[0a_{t2}\ldots a_{tk}][b_{11}b_{12}\ldots b_{1,k–1}]\ldots[b_{r-t,1}b_{r2}\ldots b_{r-t,k–1}]0_1 0_2\ldots 0_s.$$

However, under the inhomogeneous group transformations (A-3) with $\widetilde{m}_{ab} = \delta_{ab}$, $\widetilde{\ell}_a \neq 0$, we have

$$\begin{aligned}
&[0a_{12}\ldots a_{1k}][0a_{22}\ldots a_{2k}]\ldots[0a_{t2}\ldots a_{tk}][b_{11}b_{12}\ldots b_{1,k–1}]\ldots[b_{r-t,1}b_{r2}\ldots b_{r-t,k–1}]0_1 0_2\ldots 0_s \\
&\qquad\qquad \longrightarrow \\
&[0a_{12}\ldots a_{1k}][0a_{22}\ldots a_{2k}]\ldots[0a_{t2}\ldots a_{tk}][b_{11}b_{12}\ldots b_{1,k–1}]\ldots[b_{r-t,1}b_{r2}\ldots b_{r-t,k–1}]0_1 0_2\ldots 0_s + \\
&\widetilde{\ell}_{a_{12}}[00\ldots a_{1k}][0a_{22}\ldots a_{2k}]\ldots[0a_{t2}\ldots a_{tk}][b_{11}b_{12}\ldots b_{1,k–1}]\ldots[b_{r-t,1}b_{r2}\ldots b_{r-t,k–1}]0_1 0_2\ldots 0_s + \ldots \\
&+ \widetilde{\ell}_{b_{11}}[0a_{12}\ldots a_{1k}][0a_{22}\ldots a_{2k}]\ldots[0a_{t2}\ldots a_{tk}][0b_{12}\ldots b_{1k–1}]\ldots[b_{r-t,1}b_{r2}\ldots b_{r-t,k–1}]0_1 0_2\ldots 0_s \\
&+ \widetilde{\ell}_{b_{12}}[0a_{12}\ldots a_{1k}][0a_{22}\ldots a_{2k}]\ldots[0a_{t2}\ldots a_{tk}][b_{11}0\ldots b_{1,k–1}]\ldots[b_{r-t,1}b_{r2}\ldots b_{r-t,k–1}]0_1 0_2\ldots 0_s \\
&\qquad\qquad\qquad\qquad + \ldots,
\end{aligned}$$

wherein the coefficients of the $\widetilde{\ell}_{a\dots}$ terms vanish by anti-symmetry, but those of the $\widetilde{\ell}_{b\dots}$ terms clearly do not. The components corresponding to the desired $\overline{\lambda} = (r^{k-1})$ one-dimensional representation of $GL(k-1)$ within $\lambda = (r+s, r^{k-2}, t)$ is therefore *not* invariant under inhomogeneous transformations corresponding to translations in $GL(k-1) \ltimes T(k-1) \cong GL_1(k)$ *unless* the $[b_{1,1} \dots b_{r-t,k-1}]$ columns are absent, that is, $t \equiv r$. Thus, the requirement of invariance of the one-dimensional representations under $GL_1(k)$ necessitates $\lambda = (r+s, r^{k-1})$ as claimed in part (ii) of Theorem 1.

B The construction of Markov invariants

The standard construction of the irreducible modules V^{λ} is given, for example, in [24, Lecture 4]. Here we modify this procedure, to give the explicit polynomial form of the Markov invariants.

Consider the representation of $\mathfrak{S}_m$ on $\otimes^m V$ defined by the action $v_1 \otimes v_2 \otimes \dots v_m \mapsto v_{\alpha(1)} \otimes v_{\alpha(2)} \otimes \dots \otimes v_{\alpha(m)}$ for all $\alpha \in \mathfrak{S}_m$. Given a *standard* tableau $\mathfrak{T}$ with shape λ and $|\lambda| = m$, define the permutations $p \in \mathfrak{S}_m$ as those that interchange the integers in the same row, and the permutations $q \in \mathfrak{S}_m$ as those that interchange numbers in the same column. In the algebra of the representation of the symmetric group whose action is defined above, consider the quantities

$$A = \sum_{p \in \mathfrak{T}} p,$$

and

$$B = \sum_{q \in \mathfrak{T}} \text{sign}(q) q.$$

The Young's operator corresponding to $\mathfrak{T}$ is then defined as

$$Y^{\lambda} = BA.$$

It follows that for a standard tableau of shape λ, the corresponding Young's operator projects onto an irreducible module of $GL(k)$:

$$V^{\lambda} = Y^{\lambda} \cdot \otimes^m V.$$

This construction is independent of k, and Young's operators corresponding to standard tableau of the same shape project onto equivalent modules. The independent tensor components of these irreducible modules are found by inserting integers from *semi-standard* tableaux into the indices of the generic tensor. To compute the explicit form of Markov invariants, we must apply this standard procedure to our special case.

Begin with the generic form of a monomial in the components of the tensor $\psi \in \otimes^m V$:

$$\psi_{i_1 \dots i_m} \psi_{i_{m+1} \dots i_{2m}} \dots \psi_{i_{m(d-1)+1} \dots i_{md}}.$$

To find the polynomial form of an invariant that arises from an inner product of Schur functions $\{\sigma_1\} \odot \{\sigma_2\} \odot \dots \odot \{\sigma_m\}$ with $\sigma_a = \{r+s, r^{k-1}\}$ for all $1 \leq a \leq m$, and $rk + s = d$, we must apply the Young's operators to these indices. In an abuse of notation we write

$$\Psi_{i_1 \dots i_{dm}} := Y^{\sigma_1} Y^{\sigma_2} \dots Y^{\sigma_m} \psi_{i_1 \dots i_m} \psi_{i_{m+1} \dots i_{2m}} \dots \psi_{i_{m(d-1)+1} \dots i_{md}},$$

where each Young's operator Y^{σ_a}, $1 \leq a \leq m$, is generated from a standard tableau of shape $\{\sigma_a\}$ with integers chosen from the set $\{a, m+a, \dots, (d-1)m + a\}$. That is, each Y^{σ_a} permutes the indices $i_a, i_{a+m}, \dots, i_{a+md}$. The final step is to insert indices into Ψ using the semi-standard tableau which results from filling the 1st row with the integer 0, and, for $2 \leq i \leq k$, the ith row with the integer i. The justification for filling the "overhang" of length s in the first row of the tableau with the integer 0, is that in the basis given in Appendix A, the $\widetilde{\delta_0}$ component is an invariant subspace. For more details, including multiple examples, see [81].

This procedure has been implemented to garner the polynomial form of the Markov invariants for phylogenetic trees with up to four leaves. These are presented in Table 2. The algorithms required were performed in `Mathematica` [90], and, unfortunately, do not scale well for trees with more leaves. We are currently investigating the design of efficient algorithms for this construction, and note here that [64] provides a promising direction.

Additionally, in this construction the resulting polynomial form of the invariant is not in the $\delta_1, \delta_2, \ldots, \delta_k$ basis, and the required change of basis computation has thus far not been feasible. To evaluate the invariants on observed data, we therefore proceed by transforming the data itself into the appropriate basis. This allows us to evaluate Markov invariants on observed character pattern counts taken from phylogenetic data sets.

The calculation of unbiased forms, as defined in §2.2, is straight forward in principle. However, the calculation requires that the invariants be expressed in the $\delta_1, \delta_2, \ldots, \delta_k$ basis. This appears to be a rather challenging computational task, and to date the required algorithms have not been developed.

C Kimura 3ST model and phylogenetic invariants

Our approach to phylogenetic models via group actions and representations finds specific application in some special cases, such as the Kimura 3ST [49] model and certain generalizations to be described below. Here we provide a brief discussion as an illustration of our focus.

In the usual basis of point measures $\delta_A, \delta_C, \delta_U, \delta_G$, the K3ST rate matrix Q,

$$\begin{bmatrix} Q_{AA} & Q_{AG} & Q_{AU} & Q_{AC} \\ Q_{GA} & Q_{GG} & Q_{GU} & Q_{GC} \\ Q_{UA} & Q_{UG} & Q_{UU} & Q_{UC} \\ Q_{CA} & Q_{CG} & Q_{CU} & Q_{CC} \end{bmatrix} = -(\alpha{+}\beta{+}\gamma)1 + \begin{bmatrix} 0 & \alpha & \beta & \gamma \\ \alpha & 0 & \gamma & \beta \\ \beta & \gamma & 0 & \alpha \\ \gamma & \beta & \alpha & 0 \end{bmatrix} \tag{B-1}$$

can be re-written [5],

$$Q = (\alpha{+}\beta{+}\gamma)\left(-1 + \frac{\alpha}{\alpha+\beta+\gamma}K_\alpha + \frac{\beta}{\alpha+\beta+\gamma}K_\beta + \frac{\gamma}{\alpha+\beta+\gamma}K_\gamma\right), \tag{B-2}$$

where the three 'Kimura matrices'

$$K_\alpha = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad K_\beta = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad K_\gamma = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \tag{B-3}$$

span a Cartan (maximal commuting) sub*algebra* of the group $SL(4)$, and therefore can be diagonalised simultaneously, via the well-known Hadamard transform [31],

$$\mathsf{H} = h{\otimes}h{=}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \quad , \quad \mathsf{H}K_\alpha\mathsf{H}^{-1}{=}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix},$$

$$,$$

$$\mathsf{H}K_\beta\mathsf{H}^{-1}{=}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad , \quad \mathsf{H}K_\gamma\mathsf{H}^{-1}{=}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

$$\text{with} \qquad \mathsf{h} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{B-4}$$

This simple observation means that under this model, rank-m phylogenetic tensors have a spectral resolution given directly in terms of weights of the appropriate $\times^m(gl(1)\times gl(1)\times gl(1))$ abelian subalgebra of $\times^m GL(4)$ (equivalently the weight decomposition of the corresponding representation of $\times^m SL(4)$).

In fact, a stronger statement is possible. The action of group elements of the form $M(t)=e^{tQ}$ turns out to be covariant with respect to the operator δ introduced in §2.4 above, describing branching in the general phylogenetic model – explicitly, in the notation of §2.4, we have

$$\delta\cdot\exp(aK_\alpha+bK_\beta+cK_\gamma)=\exp(aK_\alpha\otimes K_\alpha+bK_\beta\otimes K_\beta+cK_\gamma\otimes K_\gamma)\cdot\delta. \tag{B-5}$$

Applied to a phylogenetic tensor P with underlying arbitrary tree $\mathcal{T}$, (B-5) then means that, under this model, the action of the Markov operators on each internal edge can be pulled back to the pendant edges, at the expense of a more complicated edge-mixing transformation. In final form, P is given by the action of a certain element of $(GL(1)\times GL(1)\times GL(1))^{\times m}$ within $GL(4^m)$, with the embedding fixed by the tree, applied to the maximally branched product measure $\delta^{(m-1)}\cdot\pi$, defined by

$$\delta^{(m-1)}\cdot\pi=\sum_i\pi_i\delta_i\otimes\cdots\otimes\delta_i,$$

with m tensor products in each term[12]. Further details can be found in [5].

Analyses of this sort are useful both analytically, and in explicit calculations. In particular, the identification of phylogenetic invariants for given trees becomes straightforward, once the components of P are written in the diagonal Hadamard basis. The group representation analysis provides a useful alternative to discrete Fourier transform methods which have been successfully applied where rate matrices admit a symmetry with respect to a discrete colour group, $Z_2\times Z_2\times\cdots$ [31, 32], and may also be useful in the characterisation of phylogenetic varieties in the phylogenetic invariants analysis [2, 80] (see also the discussion in §5).

The above considerations generalize to the case of any k-state model wherein the off-diagonal part of the rate operator is a linear combination of a maximal set of commuting permutation matrices belonging to $\mathfrak{S}_k$, which guarantees (B-5). For example, this class would include a 3-state model even simpler than the K3ST model, but which is non-symmetric, and for which the Hadamard basis is complex:

$$\begin{aligned}Q&=(\alpha{+}\beta)\big(-1+\frac{\alpha}{\alpha+\beta}K_\alpha+\frac{\beta}{\alpha+\beta}K_\beta\big),\\ K_\alpha&=\begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix},\qquad K_\beta=\begin{bmatrix}0&1&0\\0&0&1\\1&0&0\end{bmatrix}.\end{aligned}\tag{B-6}$$

See [5] for further details

References

[1] E. S. Allman and J. A. Rhodes. Phylogenetic invariants of the general Markov model of sequence mutation. *Math. Biosci.*, 186:113–144, 2003.

[2] E. S. Allman and J. A. Rhodes. Phylogenetic ideals and varieties for the general Markov model. *Adv. Appl. Math.*, to appear, 2007.

[3] A. Baker. *Matrix Groups: An Introduction to Lie Group Theory.* Springer-Verlag, 2003.

[4] D. Barry and J. A. Hartigan. Asynchronous distance between homologous DNA sequences. *Biometrics*, 43:261–276, 1987.

[12] This construction can be achieved by noting, for any linear operators A,B,C,D with $AC{=}CA$ and $BD{=}DB$, algebraic identities like $e^A\otimes e^B=e^{(A\otimes 1+1\otimes B)}$, and $e^A\otimes e^B\cdot e^{C\otimes D}=e^{(A\otimes 1+1\otimes B+C\otimes D)}$.

[5] J. D. Bashford, P. D. Jarvis, J. G. Sumner, and M. A. Steel. $U(1) \times U(1) \times U(1)$ symmetry of the Kimura 3ST model and phylogenetic branching processes. *J. Phys. A Math. Gen.*, 37:L1–L9, 2004.

[6] L. C. Biedenharn and J. D. Louck. *The Racah-Wigner Algebra in Quantum Theory.* Addison-Wesley, 1981.

[7] L. C. Biedenharn and J. D. Louck. Inhomogeneous basis set of symmetric polynomials defined by tableaux. *Proc. Natl. Acad. Sci. U.S.A.*, 87:1441–1445, 1990.

[8] O. R. P. Bininda-Emonds, editor. *Phylogenetic Supertrees: Combining Information to Reveal the Tree of Life.* Springer, 2004.

[9] D. Bryant. On the uniqueness of the selection criterion in Neighbor-Joining. *J. Class.*, 22:3–15, 2005.

[10] D. Bryant, N. Galtier, and M.-A. Poursat. Likelihood calculation in molecular phylogenetics. In Olivier Gascuel, editor, *Mathematics of Evolution and Phylogenetics*, pages 33–62. Oxford University Press, 2005.

[11] K. P. Burnham and D. Anderson. *Model Selection and Multi-Model Inference.* Springer-Verlag, 2002.

[12] M. Casanellas and J. Fernández-Sánchez. Performance of a new invariants method on homogeneous and nonhomogeneous quartet trees. *Mol. Biol. Evol.*, 24:288–293, 2007.

[13] J. A. Cavender and J. Felsenstein. Invariants of phylogenies in a simple case with discrete states. *J. Class.*, 4:57–71, 1987.

[14] M. A. Charleston. Hitch-hiking: A parallel heuristic search strategy, applied to the phylogeny problem. *J. Comput. Biol.*, 8:79–91, 2001.

[15] V. Coffman, J. Kundu, and W. K. Wootters. Distributed entanglement. *Phys. Rev. A*, 61(5):052306, Apr 2000.

[16] W. Dur, G. Vidal, and J. I. Cirac. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A*, 62:062314, 2000.

[17] N. Eriksson. Using invariants for phylogenetic tree construction. *eprint arXiv:0709.2890*, to appear.

[18] N. Eriksson and Y. Yao. Metric learning for phylogenetic invariants. *eprint arXiv:q-bio/0703034*, 2008.

[19] S. N. Evans and T. P. Speed. Invariants of some probability models used in phylogenetic inference. *Annals of Statististics*, 21(1):355–377, 1993.

[20] B. Fauser, P. D. Jarvis, R. C. King, and B. G. Wybourne. New branching rules induced by plethysm. *J. Phys. A Math. Gen.*, 39:2611–2655, 2006.

[21] J. Felsenstein. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.*, 27:401–410, 1978.

[22] J. Felsenstein. *Inferring Phylogenies.* Sinauer Associates, 2004.

[23] J. Felsenstein. *PHYLIP (Phylogeny Inference Package) version 3.6.* Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, 2005.

[24] W. Fulton and J. Harris. *Representation Theory.* Graduate Text in Mathematics. Springer-Verlag, 1991.

[25] O. Gascuel, editor. *Mathematics of Evolution and Phylogenetics.* Oxford University Press, 2005.

[26] O. Gascuel and M. Steel. Neighbor-Joining revealed. *Mol. Biol. Evol.*, 23:1997–2000, 2006.

[27] G. S. Goodman. An intrinsic time for non-stationary finite Markov chains. *Probab. Theor. Relat. Field.*, 16:165–180, 1970.

[28] R. Goodman and N. R. Wallach. *Representations and Invariants of the Classical Groups.* Cambridge University Press, 1998.

[29] X. Gu and W. H. Li. Bias-corrected paralinear and logdet distances and tests of molecular clocks and phylogenies under non-stationary nucleotide frequencies. *Mol. Biol. Evol.*, 13:1375–1383, 1996.

[30] P. R. Halmos. *Measure Theory.* Springer-Verlag, 1974.

[31] M. D. Hendy and D. Penny. Spectral analysis of phylogenetic data. *J. Class.*, 10:1–20, 1993.

[32] M. D. Hendy, D. Penny, and M. Steel. A discrete Fourier analysis for evolutionary trees. *Proc. Natl. Acad. Sci.*, 91:3339–3343, 1994.

[33] D. Hillis, J. Huelsenbeck, and D. Swofford. Hobgoblin of phylogenetics? *Nature*, 369:363–364, 1994.

[34] W. Hordijk and O. Gascuel. Improving the efficiency of SPR moves in phylogenetic tree search methods based on maximum likelihood. *Bioinformatics*, 21:4338–4347, 2005.

[35] J. P. Huelsenbeck. Performance of phylogenetic methods in simulation. *Syst. Biol.*, 44:17–48, 1995.

[36] J. P. Huelsenbeck, B. Larget, and M. E. Alfaro. Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol. Biol. Evol.*, 21:1123–1133, 2004.

[37] Daniel H. Huson, Scott M. Nettles, and Tandy J. Warnow. Disk-covering, a fast-converging method for phylogenetic tree reconstruction. *J. Comput. Biol.*, 6:369–386, 1999.

[38] M. Iosifescu. *Finite Markov Processes and Their Applications.* John Wiley and Sons, Chichester, 1980.

[39] C. Itzykson and J-B. Zuber. *Quantum Field Theory.* McGraw-Hill, New York, 1980.

[40] P. D. Jarvis, J. D. Bashford, and J. G. Sumner. Path integral formulation and Feynman rules for phylogenetic branching models. *J. Phys. A Math. Gen.*, 38:9621–9647, 2005.

[41] V. Jayaswal, L. S. Jermiin, and J. Robinson. Estimation of phylogeny using a general Markov model. *Evolutionary Bioinformatics Online*, 1:62–80, 2005.

[42] V. Jayaswal, J. Robinson, and L. Jermiin. Estimation of phylogeny and invariant sites under the general Markov model of nucleotide sequence evolution. *Syst. Biol.*, 56:155–162, 2007.

[43] L. S. Jermiin, S. Y. W. Ho, F. Ababneh, J. Robinson, and A. W. D. Larkum. Hetero: A program to simulate the evolution of DNA on four-taxon trees. *Appl. Bioinformatics*, 2:159–163, 2003.

[44] L. S. Jermiin, S. Y. W. Ho, F. Ababneh, J. Robinson, and A. W. D. Larkum. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Syst. Biol.*, 53:638–643, 2004.

[45] L. S. Jermiin, V. Jayaswal, F. Ababneh, and J. Robinson. Phylogenetic model evaluation. In J. Keith, editor, *Bioinformatics - Volume I: Data, Sequences Analysis and Evolution*, pages 331–363. Humana Press, Totowa, NJ, 2008.

[46] J. E. Johnson. Markov-type Lie groups in $GL(n, R)$. *J. Math. Phys.*, 26:252–257, 1985.

[47] A. Kelarev. *Ring Constructions and Applications.* World Scientific, 2002.

[48] R. Keown. *An Introduction to Group Representation Theory.* Academic Press, New York, 1975.

[49] M. Kimura. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci.*, 78:1454–1458, 1981.

[50] R. C. King. Branching rules for classical Lie groups using tensor and spinor methods. *J. Phys. A Math. Gen.*, 8:429–449, 1975.

[51] J. A. Lake. A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. *Mol. Biol. Evol.*, 4:167–191, 1987.

[52] J. A. Lake. Reconstructing evolutionary trees from DNA and protein sequences: Paralinear distances. *Proceedings of the National Academy of Sciences*, 91:1455–1459, 1994.

[53] J. M. Landsberg and L. Manivel. Generalizations of Strassens equations for secant varieties of Segre varieties. *Communications in Algebra*, 36:405–422, 2008.

[54] D. E. Littlewood. *The Theory of Group Characters.* Clarendon Press, Oxford, 1940.

[55] D. E. Littlewood. Plethysm and the inner product of S-functions. *J. Lond. Math. Soc.*, s1–32:18–22, 1955.

[56] P. J. Lockhart, A. W. D. Larkum, M. A. Steel, P. J. Waddell, and D. Penny. Evolution of chlorophyll and bacteriochlorophyll: The problem of invariant sites in sequence analysis. *Proc. Natl. Acad. Sci. U.S.A.*, 93:1930–1943, 1996.

[57] P. J. Lockhart, P. Novis, B. G. Milligan, J. Riden, A. Rambaut, and A. W. D. Larkum. Heterotachy and tree building: A case study with plastids and eubacteria. *Mol. Biol. Evol.*, pages 40–45, 2006.

[58] P. J. Lockhart, M. A. Steel, A. C. Barbrook, D. H. Huson, and C. J. Howe. A covariotide model describes the evolution of oxygenic photosynthesis. *Mol. Biol. Evol.*, 15:1183–1188, 1998.

[59] P. J. Lockhart, M. A. Steel, M. D. Hendy, and D. Penny. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.*, 11:605–612, 1994.

[60] I. G. MacDonald. *Symmetric Functions and Hall Polynomials.* Clarendon Press, Oxford, 1979.

[61] T. Massingham and N. Goldman. Statistics of the log-det estimator. *MBE Advance Access published August 16, 2007*, 2007.

[62] F. A. Matsen and S. N. Evans. Ubiquity of synonymity: Almost all large binary trees are not uniquely identified by their spectra or their immanantal polynomials. *eprint arXiv:q-bio/0512010*, 2006.

[63] P. McCullagh. *Tensor Methods in Statistics.* Chapman and Hall, 1987.

[64] A. Molev. On the fusion procedure for the symmetric group. *eprint arXiv:math/0612207*, 2007.

[65] B. Mourad. On a Lie-theoretic approach to generalised doubly stochastic matrices and applications. *Linear and Multilinear algebra*, 52:99–113, 2004.

[66] P. J. Olver. *Classical Invariant Theory.* Cambridge University Press, Cambridge, 2003.

[67] M. Pagel and A. Meade. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. *Syst. Biol.*, 53:571–581, 2004.

[68] E. Paradis, J. Claude, and K. Strimmer. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20:289–290, 2004.

[69] D. Penny, B. J. McComish, M. A. Charleston, and M. D. Hendy. Mathematical elegance with biochemical realism: the covarion model of molecular evolution. *J. Mol. Evol.*, 53:711–723, 2001.

[70] D. Posada and T. R. Buckley. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.*, 53:793–808, 2004.

[71] R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2006.

[72] C. Semple and M. Steel. *Phylogenetics.* Oxford Press, 2003.

[73] C. S. Seshadri. On a theorem of Weitzenböck in invariant theory. *J. Math. Kyoto. Univ.*, 1:403–409, 1962.

[74] M. Steel. Some statistical aspects of the maximum parsimony method. In R. DeSalle, G. Giribet, and W. Wheeler, editors, *Molecular Systematics and Evolution: Theory and Practice*, pages 125–140. Birkhäuser Verlag, 2002.

[75] M. Steel. Should phylogenetic models be trying to fit an elephant? *Genetics*, 21:307–309, 2005.

[76] M. A. Steel. Recovering a tree from the leaf colourations it generates under a Markov model. *Appl. Math. Lett.*, 7:19–24, 1994.

[77] M. A. Steel, L. Szekely, P. L. Erdos, and P. Waddell. A complete family of phylogenetic invariants for any number of taxa under Kimura's 3ST model. *N.Z. J. Bot.*, 31:289–296, 1993.

[78] K. Strimmer and A. von Haeseler. Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.*, 13:964–960, 1996.

[79] B. Sturmfels. Open problems in algebraic statistics. In *M. Putinar and S. Sullivant (Eds.), Emerging Applications of Algebraic Geometry, I.M.A. Volumes in Mathematics and its Applications*, to appear.

[80] B. Sturmfels and S. Sullivant. Toric ideals of phylogenetic invariants. *J. Comput. Biol.*, 12:204–228, 2005.

[81] J. G. Sumner. Entanglement, Invariants, and Phylogenetics. *PhD thesis, University of Tasmania,* *`http://eprints.utas.edu.au`*, 2006.

[82] J. G. Sumner. Phylogenetic quartet inference using the squangles. *University of Sydney,* *`http://www.it.usyd.edu.au/~mcharles/software`*, 2008.

[83] J. G. Sumner and P. D. Jarvis. Entanglement invariants and phylogenetic branching. *J. Math. Biol.*, 51:18–36, 2005.

[84] J. G. Sumner and P. D. Jarvis. Using the tangle: A consistent construction of phylogenetic distance matrices. *Math. Biosci.*, 204:49–67, 2006.

[85] C. Tuffley and M. A. Steel. Links between maximum likelihood and maximum parsimony under a simple model of site substitution. *Bull. Math. Biol.*, 59:581–607, 1997.

[86] R. Weitzenböck. Über die Invarianten von linearen Gruppen. *Acta. Math.*, 58:231–293, 1931.

[87] H. Weyl. *The Theory of Groups and Quantum Mechanics.* Dover Publications, 1950.

[88] M. L. Whippman. Branching rules for simple Lie groups. *J. Math. Phys.*, 6:1534–1539, 1965.

[89] M. Wilkinson and J. A. Cotton. Supertree methods for building the tree of life: Divide-and-conquer approaches to large phylogenetic problems. In T. R. Hodkinson and J. A. N. Parnell, editors, *Reconstructing the Tree of Life: Taxonomy and Systematics of Species Rich Taxa. Systematics Association Special Volume 72.* CRC Press, 2006.

[90] Wolfram Research, Inc. `Mathematica 5.2`. 2005.

[91] B. G. Wybourne. `Schur`: An interactive programme for calculating properties of Lie groups. version 6.03. *http://sourceforge.net/projects/schur*, 2004.

[92] Z. Yang. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.*, 39:306–314, 1994.

[93] Z. Yang. *Computational Molecular Evolution.* Oxford University Press, 2006.

[94] A. Zharkikh. Estimation of evolutionary distance between nucleotide sequences. *J. Mol. Evol.*, 39:315–329, 1994.